

FUNCTIONS OF ORGANELLE-SPECIFIC NUCLEIC ACID  
BINDING PROTEIN FAMILIES IN CHLOROPLAST  
GENE EXPRESSION

by

JANA PRIKRYL

A DISSERTATION

Presented to the Department of Biology  
and the Graduate School of the University of Oregon  
in partial fulfillment of the requirements  
for the degree of  
Doctor of Philosophy

December 2009

**University of Oregon Graduate School**

**Confirmation of Approval and Acceptance of Dissertation prepared by:**

Jana Prikryl

Title:

"Functions of Organelle-Specific Nucleic Acid Binding Protein Families in Chloroplast Gene Expression"

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Biology by:

Eric Selker, Chairperson, Biology  
Alice Barkan, Advisor, Biology  
Victoria Herman, Member, Biology  
Karen Guillemin, Member, Biology  
J. Andrew Berglund, Outside Member, Chemistry

and Richard Linton, Vice President for Research and Graduate Studies/Dean of the Graduate School for the University of Oregon.

December 12, 2009

Original approval signatures are on file with the Graduate School and the University of Oregon Libraries.



with PPR5 and PPR10 suggest that most of these activities may result directly from the unusually long RNA binding surface predicted for PPR proteins, which we have shown imparts two biochemical properties: site-specific protection of RNA from other proteins and site-specific RNA unfolding activity. I narrowed down the binding site for PPR5 and PPR10 to ~45 nt and 19 nt, respectively. I showed that PPR5 contributes to the splicing of its group II intron ligand by restructuring sequences that are important for splicing. I used *in vitro* assays with purified PPR10 to confirm that PPR10 can block exonucleolytic RNA decay from both the 5' and 3' directions, as predicted by prior *in vivo* data. I also present evidence that PPR10 promotes translation by restructuring its RNA ligand to allow access to the ribosome. These findings illustrate how the unusually long RNA interaction surface predicted for PPR proteins can have diverse effects on RNA metabolism.

This dissertation includes both previously published and unpublished co-authored material.

## CURRICULUM VITAE

NAME OF AUTHOR: Jana Prikryl

PLACE OF BIRTH: Olomouc, Czech Republic

DATE OF BIRTH: March 30, 1976

## GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene  
University of Colorado, Boulder  
University of Colorado, Colorado Springs  
Pikes Peak Community College, Colorado Springs

## DEGREES AWARDED:

Doctor of Philosophy, Department of Biology, 2009, University of Oregon  
Bachelor of Arts, 1998, University of Colorado, Boulder

## AREAS OF SPECIAL INTEREST:

Molecular and Cellular Biology  
Genetics  
Biochemistry

## PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellow, Department of Biology, University of Oregon,  
Eugene, 2004-2005 and 2008-2009

Teaching Assistant, Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, 2002-2004

Professional research assistant, lab of Professor Kathleen Danna, University of Colorado, Boulder, 1998-2001

Professional research assistant, lab of Professor Peter Kuempel, University of Colorado, Boulder, 1998-2000

#### GRANTS, AWARDS AND HONORS:

American Association for the Advancement of Science (AAAS)/ Program for Excellence in Science, 2008-2009

Genetics training grant, National Institute of Health (NIH), 2004-2006 and 2008

Molecular Biology training grant, National Institute of Health (NIH), 2006-2007

Co-president of Students in Biological Sciences (SIBS) graduate student group, 2006-2007

Undergraduate Research Opportunities (UROP) Grant, University of Colorado, 1998

#### PUBLICATIONS:

Pfalz J, Ali Bayraktar O, Prikryl J, and Barkan A. (2009) Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO J* 28(14):2042-52.

Prikryl J, Watkins KP, Friso G, van Wijk KJ, Barkan A. (2008) A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis. *Nucleic Acids Res* 36(16):5152-65.

Prikryl J, Hendricks EC, Kuempel PL. (2001) DNA degradation in the terminus region of resolvase mutants of *Escherichia coli*, and suppression of this degradation and the Dif phenotype by recD. *Biochimie* 83(2):171-6.

## ACKNOWLEDGMENTS

I am sincerely grateful to my advisor Alice Barkan for her encouragement, support and thoughtfulness. She has gone far beyond the call of duty to help me progress as a scientist. Sincere and conscientious, she is not only a wonderful mentor but also, a model of what I believe a scientist, collaborator, and educator should be. I am extremely thankful for my lab mates Kenneth Watkins, Tiffany Kroeger, Margarita Rojas, Rosalind Williams-Carrier, Susan Belcher, Yukari Asakura, and Jeannette Pfalz. They are invaluable to me as both coworkers and as friends. They have always made time to give me helpful advice and share their vast knowledge. They have supported me through failure and success. They have forgiven me for being irritable at times. Their camaraderie and passion for their work is uplifting and inspiring.

I am very appreciative to the members of my committee Eric Selker, Tory Herman, Karen Guillemin, and Andy Berglund. Their helpful advice and involvement in the progress of my graduate career have truly made a positive impact. I am thankful to the staff of the Institute of Molecular Biology and the Biology Department, they have made my life easier in so many ways I cannot list them all here. I am ever grateful for my friends Kohl, Bob, Tiffany, Luke, Emily, Scott, Clair, and Andy for their companionship and support, and to my parents Jarmila and Ivan and my sister Helena who are the model of courage, and love, and who give me strength to move forward.

## DEDICATION

This dissertation is dedicated to Professor Nancy Guild whose kind nature, devotion to teaching, and classroom ingenuity continue to inspire me, and whose encouragement and support gave me the confidence to set out on this path.



## TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION .....	1
Co-Evolution of the Chloroplast and Nuclear Genomes .....	1
RNA Metabolism in the Chloroplast .....	2
Non-Canonical RNA Binding Proteins in the Organelles .....	3
II. A MEMBER OF THE WHIRLY FAMILY IS A MULTIFUNCTIONAL RNA AND DNA BINDING PROTEIN THAT IS ESSENTIAL FOR CHLOROPLAST BIOGENESIS .....	5
Introduction .....	5
Materials and Methods .....	7
Purification of CRS1 Ribonucleoproteins and Mass Spectrometry .....	7
Plant Material .....	7
Generation of Recombinant ZmWHY1 for Antibody Production and Binding Assays .....	7
Chloroplast Fractionation and Protein Analysis .....	8
Nucleic Acid Coimmunoprecipitation Assays .....	8
Analysis of DNA and RNA .....	9
Nucleic Acid Binding Assays .....	9
Chloroplast Run-On Transcription Assay .....	10
Results .....	10
Identification of ZmWHY1 in CRS1 Coimmunoprecipitates .....	10
Recovery of <i>ZmWhy1</i> Insertion Mutants .....	11
ZmWHY1 Partitions Between the Chloroplast Stroma and Thylakoid Membrane, To Which it is Bound in a DNA-Dependent Manner .....	14
ZmWHY1 Is Associated With Large RNA- and DNA-Containing Particles .....	14

Chapter	Page
Coimmunoprecipitation Assays Demonstrate That ZmWHY1 Associates with a Subset of Plastid RNAs That Includes the <i>atpF</i> Intron .....	17
DNA From Throughout the Plastid Genome Coimmunoprecipitates with ZmWHY1 .....	20
Zm <i>Why1</i> Mutants Are Deficient for Plastid Ribosomes .....	21
ZmWHY1 Promotes <i>atpF</i> Intron Splicing .....	25
ZmWHY1 Is Required Neither for Chloroplast DNA Replication nor for Global Plastid Transcription .....	27
Recombinant ZmWHY1 Binds Single-Stranded RNA and DNA <i>in Vitro</i> .....	30
Discussion .....	32
Multiple Roles of ZmWHY1 in Chloroplast Biogenesis .....	32
ZmWHY1 Binds both RNA and DNA <i>in Vitro</i> and <i>in Vivo</i> .....	33
What is WHY1's DNA-Related Function in the Chloroplast? .....	34
Bridge .....	36
III. BIOCHEMICAL ANALYSES SUGGEST THAT PPR/RNA INTERACTIONS INVOLVE AN UNUSUAL RNA/PROTEIN INTERFACE THAT IS SUFFICIENT TO MEDIATE A VARIETY OF POSTTRANSCRIPTIONAL EFFECTS .....	37
Introduction .....	37
Materials and Methods .....	39
Ribonucleic Acid Binding Assays .....	39
Minimal Binding Assay Using Partially Alkali Hydrolyzed RNA .....	40
PNPase Purification .....	40
<i>In Vitro</i> Exonuclease Protection Assays .....	41
Nuclease Cleavage Structure Probing Assays .....	42
2-Aminopurine Fluorescence Assay .....	43
Results .....	43
The Minimal PPR10 Binding Site Spans 15 Nucleotides .....	43
PPR10 Protects its RNA Ligand From 3' and 5' Exonucleolytic Cleavage <i>in Vitro</i> .....	44

Chapter	Page
PPR10 Binding Releases the <i>atpH</i> Ribosome Binding Site From a Sequestering Secondary Structure .....	49
The PPR5 Binding Site is Complex and Includes Discontinuous RNA Segments .....	51
PPR5-Induced Changes in RNA Structure Suggest Mechanisms by which PPR5 Enhances Splicing .....	56
Discussion .....	62
Features of the PPR10 Binding Site Suggest That PPR10 Binds RNA Along an Unusually Long RNA/Protein Interface .....	62
Site-Specific Barrier and RNA Remodeling Functions of PPR5 and PPR10: Implications for the Mechanisms by which PPR Proteins Mediate Downstream Effects .....	65
IV. CONCLUSIONS AND FUTURE DIRECTIONS .....	69
Conclusions .....	69
Future Directions .....	72
Future Directions Related to WHY1 .....	72
Immediate Directions Related to PPR Proteins .....	72
Long-Term Directions Related to PPR Proteins .....	73
REFERENCES .....	76

## LIST OF FIGURES

Figure	Page
CHAPTER II	
1. Mutant Alleles of <i>ZmWhy1</i> .....	12
2. Intracellular Localization of ZmWHY1 .....	15
3. Sucrose-gradient sedimentation demonstrating that ZmWHY1 is associated with DNA- and RNA-containing particles in chloroplast stroma .....	16
4. Identification of chloroplast RNAs and DNAs that coimmunoprecipitate with ZmWHY1 .....	18
5. Plastid ribosome deficiency in <i>ZmWhy1</i> mutants .....	22
6. Reduced <i>atpF</i> intron splicing in <i>ZmWhy1</i> mutants .....	26
7. Accumulation of plastid RNAs in <i>ZmWhy1</i> mutants .....	28
8. Chloroplast DNA levels in <i>ZmWhy1</i> mutants .....	29
9. Recombinant ZmWHY1 binds single-stranded RNA and DNA .....	31
CHAPTER III	
1. The PPR10 RNA ligand .....	45
2. Stoichiometric binding assay with recombinant PPR10 .....	47
3. PPR10 protects against 3' and 5' exonuclease activity <i>in vitro</i> .....	48
4. PPR10 binding induces structural changes in the <i>atpH</i> 5'UTR .....	50
5. Mapping the boundaries of sequences required for a high-affinity interaction with PPR5 .....	53
6. The PPR5 RNA ligand .....	55
7. Ribonuclease sensitivity assay of RNA structure in the absence and presence of PPR5 .....	58
8. PPR5 causes an increase in 2-aminopurine fluorescence in its ligand, indicating unfolding of the RNA stem .....	61

## CHAPTER I

### INTRODUCTION

#### **Co-Evolution of the Chloroplast and Nuclear Genomes**

The overarching goal of my graduate work has been to gain better insight into how chloroplast gene expression is regulated by the nuclear genome. In order to appreciate this process one must first quickly review the evolution of the organelles (1-3). The mitochondrion and chloroplast each arose as a result of an endosymbiotic event. First the endosymbiosis of a proteobacterium gave rise to the mitochondrion; there is consensus that this was a very early event during the evolution of the eukaryotic cell, although there is controversy concerning whether it predated the evolution of the nucleus. More recently (~1 billion years ago), the engulfment of a cyanobacterium gave rise to the chloroplast. As a consequence of these events, genetic material is found in three places in the plant cell: the nucleus, the mitochondrion, and the chloroplast. The genomes of the organelles are greatly diminished in comparison to those of their bacterial ancestors. This has occurred through successive gene loss from the organelles, sometimes in conjunction with gene transfer to the nuclear genome. Despite this gene loss, approximately 100 genes have been retained in the chloroplast genome in land plants. Most of these encode proteins that are directly involved in either photosynthesis (e.g. subunits of photosystem II, photosystem I, etc) or in chloroplast gene expression (e.g. tRNAs, rRNAs, ribosomal proteins, and RNA polymerase subunits). Because some genes encoding subunits of the photosynthetic complexes are found in the chloroplast, whereas others are encoded in the nucleus, concerted expression of the chloroplast and nuclear genomes is required for proper chloroplast function. The basis for the retention of some genes in the chloroplast is

not fully understood but has been proposed to facilitate redox based regulation of protein expression, wherein the photosynthetic status of the chloroplast can directly influence the expression of chloroplast genes involved in photosynthesis (4).

The co-evolution of the chloroplast and the host cell resulted in two classes of chloroplast targeted, nuclear encoded proteins: those derived from the ancestral cyanobacterium and those derived from the host genome. Most nuclear genes of cyanobacterial ancestry encode proteins that are targeted to the chloroplast, where they carry out their ancestral function; these compensate for the loss of the orthologous genes from the chloroplast. Host-derived proteins on the other hand lack relatives in bacteria; they are thought to have evolved from proteins with functions outside of the chloroplast, with subsequent co-opting to fulfill newly acquired needs of the chloroplast. In fact, the chloroplast acquired many features that are not characteristic of its cyanobacterial ancestor, and the emergence of these features seems to have been accompanied by the “invention” of several new nuclear-encoded protein families that are dedicated to these functions. Examples of this phenomenon are highlighted below.

### **RNA Metabolism in the Chloroplast**

The complexity of RNA metabolism in the chloroplast is much greater than that in cyanobacteria (5). For example, chloroplast RNAs are modified by RNA editing, terminal processing at both the 5' and 3' ends, group I and group II intron splicing, and intercistronic processing of polycistronic precursors (reviewed in 6, 7). Furthermore, most gene regulation is believed to occur at the post-transcriptional level, via modulation of RNA stability and translation.

There are 18 introns in the maize chloroplast genome: one group I intron and 17 group II introns. These introns are classified as autocatalytic because members of both groups in other organisms have been shown to self-splice *in vitro*. However, group I and II introns in higher plant chloroplasts are incapable of self-splicing *in vivo* and require

protein cofactors to facilitate splicing. In fact, it is thought that the nuclear spliceosome evolved from the degeneration of group II introns accompanied by the co-evolution of proteins to compensate for the loss of autocatalytic RNA activity. The splicing of chloroplast group II introns requires many nucleus-encoded proteins, but these proteins are not related to spliceosomal proteins, nor to the RNA binding protein classes that function in the nuclear-cytosolic compartment or in bacteria (6, 8, 9).

RNA editing and the processing of polycistronic precursors to single gene mRNAs are also characteristic of chloroplasts, but not of their bacterial ancestors. The intercistronic processing events generate complex populations of RNAs from most chloroplast genes, reflecting the full length polycistronic precursor, various processing intermediates, and fully-processed monocistronic mRNAs. It had been speculated that this processing arises through site-specific endonucleolytic cleavages, but our laboratory's previous work, as well as results described in this thesis, suggested an entirely different mechanism to account for the complex chloroplast transcript populations.

### **Non-Canonical RNA Binding Proteins in the Organelles**

Genetic approaches have been used to identify nucleus-encoded, proteins that affect these various aspects of chloroplast RNA metabolism. A striking finding is that the vast majority of such proteins are not related to the classic RNA binding proteins found in the nuclear-cytosolic compartment. In fact, most of these proteins belong to families that are dedicated to organelle gene expression. Some examples are the CRM, DUF860, and PPR families of proteins (8-10). These proteins harbor non-canonical RNA binding domains, with most or all family members targeted to the chloroplast or mitochondrion. CRM and DUF860 proteins are involved in the splicing of many chloroplast and mitochondrial introns, whereas the PPR family plays multiple roles in the chloroplast and mitochondrion, including RNA splicing, RNA editing, translational

control, and maintaining RNA stability. CRM and DUF860 proteins are plant-specific whereas PPR proteins are found in all eukaryotes. However, there has been a large expansion of the PPR family in plants; whereas there are ~10 PPR proteins in humans and yeast, there are ~450 in land plants (11). Although these protein families play essential roles in many aspects of organellar RNA metabolism, very little is known about the mechanisms by which they exert their effects.

To understand the mechanisms by which the non-canonical RNA binding proteins characteristic of the chloroplast and mitochondrion mediate their effects, I have studied three nuclear encoded proteins that are required for chloroplast biogenesis: WHY1, PPR10, and PPR5. All three of these proteins are targeted to the chloroplast, bind to RNA with sequence specificity, and are involved in multiple steps in RNA processing. WHY1 and PPR5 facilitate group II intron splicing. PPR5 and PPR10 protect RNA from nucleases, and PPR10 also promotes translation. How WHY1 exerts its downstream effects is still a mystery. Our detailed study of PPR5 and PPR10, on the other hand, is beginning to give us insight into how members of the PPR family can mediate multiple downstream effects through one biochemical property: a long and specific RNA interaction surface.

The work on WHY1 (Chapter II) has been published and is co-authored by Jana Prikryl, Kenneth P. Watkins, Giulia Friso, Klaas J. van Wijk, and Alice Barkan. The work on PPR5 and PPR10 (Chapter III) is in preparation for publishing and will also be co-authored by Jana Prikryl, Margarita Rojas, Rosalind Williams-Carrier, Omer Ali Bayraktar, and Alice Barkan.



## **CHAPTER II**

### **A MEMBER OF THE WHIRLY FAMILY IS A MULTIFUNCTIONAL RNA AND DNA BINDING PROTEIN THAT IS ESSENTIAL FOR CHLOROPLAST BIOGENESIS**

This chapter describes the characterization of a member of the plant specific, Whirly protein family, WHY1. This work was done in collaboration with Dr. Alice Barkan, and Dr. Kenneth Watkins. In addition, Dr. Giulia Friso, and Dr. Klaas van Wijk contributed by using Mass Spectroscopy to identify the WHY1 protein. This work has been published and co-authored with the above-mentioned individuals.

#### **Introduction**

Plant mitochondria and chloroplast genomes encode ~50 and ~100 products, respectively, most of which participate in basal organellar gene expression or energy transduction. Post-transcriptional events play the dominant role in dictating gene product abundance in both organelles (reviewed in 12). In fact, the two organelles house a similar repertoire of RNA processing pathways that includes RNA editing, group II intron splicing, and endonucleolytic processing. Genetic and bioinformatic analyses suggest that many hundreds of nuclear genes encode organelle-localized nucleic acid binding proteins and influence organellar gene expression (9, 11, 13, 14), but only a small fraction of such genes has been studied.

The protein that is the focus of this study, ZmWHY1, came to our attention during our characterization of the chloroplast RNA splicing machinery. Nine nucleus-encoded

proteins that are necessary for the splicing of various subsets of the ~20 chloroplast introns in vascular plants have been reported (15-24). One of the first to be characterized, CRS1, is necessary for the splicing of the group II intron in the chloroplast *atpF* gene (15, 18), and binds specifically to that intron *in vivo* and *in vitro* (19, 20, 25). However, the large size of the particles containing CRS1 and *atpF* intron RNA *in vivo*, and the fact that CRS1 is not sufficient to promote *atpF* intron splicing *in vitro* suggested that additional proteins are involved. We therefore used mass spectrometry to identify proteins that coimmunoprecipitate with CRS1; ZmWHY1 was one such protein.

ZmWHY1 is a member of the “Whirly” protein family, whose orthologs in potato (StWHY1) and Arabidopsis (AtWHY1) were reported to be nuclear transcription factors involved in pathogen-induced transcription (26, 27). StWHY1 and AtWHY1 bind single-stranded DNA *in vitro*, and StWHY1 adopts a propeller-like structure from which the family acquired its name (26, 28). AtWHY1 has also been implicated in telomere binding and maintenance (29). Additional functions for members of the Whirly family were suggested by the fact that GFP fused to each member of the family from Arabidopsis localizes to chloroplasts or mitochondria (30). The copurification of AtWHY1 with a transcriptionally-active chloroplast DNA complex (31) and the association of AtWHY2 with mitochondrial nucleoids (32) confirmed that these proteins have organellar functions, but the nature of these functions is not known. Results presented here show that ZmWHY1 plays an essential role in the biogenesis of chloroplasts, that it is associated with DNA from throughout the chloroplast genome and that it interacts *in vivo* with a subset of chloroplast RNAs that includes the *atpF* intron. ZmWHY1 enhances *atpF* intron splicing and influences the biogenesis of the large ribosomal subunit. However, chloroplast DNA and RNAs in *ZmWhy1* mutants accumulate to levels similar to those in other mutants with plastid ribosome deficiencies of similar magnitude. These results argue that ZmWHY1 is required neither for chloroplast DNA replication nor directly for global chloroplast transcription.

## Materials and Methods

### Purification of CRS1 ribonucleoproteins and mass spectrometry

Purification of CRS1 ribonucleoprotein particles and mass spectrometry were performed as described for CAF1 and CAF2 particles in ref (21). The antibody to CRS1 was described previously (20).

### Plant material

Our collection of *Mu* transposon-induced non-photosynthetic maize mutants (<http://chloroplast.uoregon.edu>) was screened by PCR to identify insertions in ZmWHY1, using methods described in (33) and a Zm*Why1*-specific primer (5'-CGGCGGCCTTTCTGGAGGA -3') in conjunction with a *Mu* terminal inverted repeat primer (5'-GCCTCCATTTTCGTCGAATCCCG -3'). The alleles were tested for complementation by crossing phenotypically normal siblings (+/+ or +/-) from ears segregating each allele. 74 ears were recovered, 36 of which segregated chlorophyll deficient mutants. Other mutants used in this work include *iojap* (34), *hcf7* (35), and *crs1* (15). The inbred line B73 (Pioneer HiBred) was used as the source of wild type tissue for coimmunoprecipitation, sucrose gradient, and chloroplast fractionation experiments. Plants were grown in soil in a growth chamber (16h light, 24°C) / 8h dark, 19°C). Leaf tissue was harvested ~9 days after planting.

### Generation of recombinant ZmWHY1 for antibody production and binding assays

ESTs representing Zm*Why1* were identified as GenBank accessions DV170433 and DV503865; the corresponding cDNAs were obtained from the maize full-length cDNA project (<http://www.maizecna.org/>). The complete cDNA sequence was determined and has been entered in GenBank under Accession EU595664. A ZmWHY1 protein fragment (amino acids 86 to 258) with a C-terminal 6x-histidine tag was expressed in *E. coli* from pET28b (Novagen), purified by nickel affinity chromatography and used for the production of polyclonal antisera in rabbits at the University of Oregon

antibody facility. Full-length mature ZmWHY1 (i.e. lacking the transit peptide) for nucleic acid binding assays was generated by PCR amplification of its coding sequence from the cDNA (primers 5'- TATAGGATCCGCCTCCTCCCGTAAG -3' and 5'- TATAGTCGACTCACCGACGCCATTC -3'), digestion of the product with BamHI and Sall, and cloning into pMAL-TEV. Subsequent steps in expressing and purifying recombinant ZmWHY1 were as described previously for RNC1 (21).

### **Chloroplast fractionation and protein analysis**

Leaf protein extracts were prepared and analyzed as previously described (36). Chloroplast subfractions were those described by Williams and Barkan (33). For RNase and DNase treatment of thylakoid membranes, MgCl<sub>2</sub> was added to a thylakoid membrane fraction to a concentration of 15 mM. The sample was divided into three 20 µl aliquots: 1 µl RNase-free RQ1 DNase (1U/µl) (Promega), 1 µl of RNase A (1 µg/µl), or 1 µl water was added for the DNase, RNase, and mock treatments, respectively. Samples were incubated at room temperature for 30 min and then centrifuged at 4°C at 15,000 x g for 15 min. The pellet was resuspended in 10 mM Tris-HCl pH 7.5, 2 mM EDTA, 0.2 M sucrose, to a volume equivalent to that of the supernatant. The supernatant and pellet fractions were analyzed by SDS-PAGE and immunoblotting. Sucrose gradient sedimentation of stromal extract was performed as described by Jenkins and Barkan (16); aliquots of stroma were treated with either 3 units RQ1 DNase or 50 µg/ml RNase A for 30 min at room temperature prior to centrifugation. Antisera to spinach chloroplast RPL2 and MDH were generously provided by A. Subramanian (University of Arizona) and Kathy Newton (University of Missouri), respectively. The other antibodies were generated by us and described previously (37).

### **Nucleic acid coimmunoprecipitation assays**

100 µl aliquots of stromal extract (~500 µg of protein) were analyzed by RIP-chip, DIP-chip, and slot-blot hybridization using methods described in (38), except that stroma used for RIP-chip assays was treated with DNase prior to immunoprecipitation

(10 units RQ1 DNase at 37 °C for 30 min) and again after purification of nucleic acids from the immunoprecipitation. For DIP-chip assays, RNase A (100 µg/ml final concentration) was added to stroma prior to immunoprecipitation and residual RNA was removed from the recovered nucleic acids by alkali hydrolysis in 200 mM NaOH at 70 °C for 30 min.

### **Analysis of DNA and RNA**

DNA extraction from leaf tissue and Southern blot analysis were performed as previously described (39). Leaf RNA was extracted from the middle of the second leaf of 9 day old seedlings, with Tri Reagent (Molecular Research Center). RNA gel blot hybridizations were performed as previously described (36). The following PCR fragments were used as probes (residue numbers refer to GenBank accession X86563): *atpF int/ex2*, 35706-36384; *atpF int*, 36073-35233; *ndhA int*, 114941-115730; orf99, 86911-88430; *petD ex2*, 75539-75895; *petN*, 19081-19415; *psbA*, 296-1074; *rpl16 ex2*, 79519-79920; *rpl16 int*, 80002-80888; *rpoB*, 23258-24475; *rps12 trans*, 69307-69420 and 129636-129861; *rps12 int1/ex1*, 5', 68793-69460; *rps14*, 38500-39020; *rrn4.5*, 102041-102135; *rrn5*, 102180-102619; *rrn16*, 95559-96779; *rrn23*, 98332-98792; *trnA* mature, 98038-98075 + 98712-98916; *trnG mature*, 13245-13292 and 13991-14013; *trnG int* 13293-13990; *trnN*, 103066-103137; *ycf3 int2/ex3*, 43820-44873; *ycf3 int*, 44383-45116. Poisoned primer extension assays to distinguish mature from precursor RNAs were performed as previously described (18) using the following primers and dideoxynucleotide: *rrn23*, 5'- CGCAAGCCTTTCCTCTTTT -3' (ddTTP); *rpl2*, 5'- GGCCGTGCCTAAGGGCATATC -3' (ddCTP); *rps12*, 5'- GGTTTTTTGGGGTTGATAG -3' (ddCTP). Radioactive gels and blots were imaged with a phosphorimager and analyzed using ImageQuant software (GE Healthcare).

### **Nucleic acid binding assays**

Gel mobility shift assays were performed with the same substrates and procedures as described in Watkins et al. (2007) (21) except that the binding reactions contained 150

mM NaCl, 5 mM DTT, 50 µg/ml BSA, 25 mM Tris-HCl pH 7.5, 0.1 mg/ml Heparin. Filter binding assays were based on the procedure of Wong and Lohman (40) with modifications (25). The *atpF* intron RNA substrate for filter-binding assays was transcribed *in vitro* by T7 RNA from a PCR product generated with the following primers: *atpF* forward /T7 promoter, 5'-TAATACGACTCACTATAGGGATGAAAAA TGTAACCCATTCTT -3'; *atpF* reverse, 5'- AATGAAAGTAGATTATCTTGC -3'. The RNA, which included *atpF* exon 1 and the complete intron, was heated in TE to 90 °C for 2 min and then placed on ice immediately prior to its addition to binding reactions (300 mM NaCl, 5 mM DTT, 50 µg/ml BSA, and 25 mM Tris pH 7.5, 30 °C for 30 min).

### **Chloroplast run-on transcription assay**

The chloroplast run-on transcription assay was performed as described by Mullet and Klein (41-43). The radiolabeled products were hybridized to the following synthetic oligonucleotides (10 pmol/slot) that had been applied with a slot-blot manifold to a nylon membrane: *rrn16* 5'-

CCCATTGTAGCACGTGTGTCGCCAGGGCATAAGGGGCATGATGACTTGG -3',  
*rrn23* 5'- GGACTCTTGGGGAAGATCAGCCTGTTATCCCTAGAGTAACTTTTATC  
 CGA -3', *trnG* 5'- CATCTATGTCAGCTTTTCTGTCTGAATGGAACCAAAGCTCTC  
 CGCTTTCTAGATGC -3', and *CFM3* 5'- AACTCGAGCGAAAAACAGGAGGATT  
 AGTAATCTGGCGATCAGGGACTTCTGTTTCTCTGTACCGGGGAGTAGATTATG  
 ATGAACC -3'.

## **Results**

### **Identification of ZmWHY1 in CRS1 coimmunoprecipitates**

To find proteins involved in the splicing of the *atpF* intron we used mass spectrometry to identify proteins that coimmunoprecipitate with the *atpF* splicing factor CRS1. Stromal extract was initially fractionated on a sucrose gradient, and the fractions

that contained the majority of the CRS1 ribonucleoprotein particles (~600-700 kDa) were used for immunoprecipitation. The immunoprecipitated proteins were separated by SDS-PAGE, and contiguous gel slices containing proteins between ~20 and ~120 kDa were used for in-gel trypsin digests and tandem mass spectrometry. Among the proteins identified was a member of the Whirly protein family (Supplementary Table 1, Supplementary Figure 1A) (26, 28). The Whirly protein family in vascular plants includes two orthologous groups (Supplementary Figure 1B). The peptides detected in the CRS1 coimmunoprecipitate identified the protein as a member of the orthologous group designated Why1.

### Recovery of *ZmWhy1* insertion mutants

To elucidate the function of ZmWHY1 we sought insertion mutants in a reverse-genetic screen of our collection of transposon-induced non-photosynthetic maize mutants (<http://pml.uoregon.edu/>). Two mutant alleles were recovered (Figure 1): the *Zmwhy1-1* allele has a *MuDR* transposon insertion 35 bp downstream of the predicted start codon and conditions an ivory leaf phenotype; the *Zmwhy1-2* allele has a *Mu1* or *Mu1.7* insertion 38 bp upstream of the predicted start codon and conditions a pale green leaf phenotype. The heteroallelic progeny of complementation crosses (*Zmwhy1-1/-2*) exhibit an intermediate phenotype (Figure 1B). Homozygous mutant plants die after the development of three to four leaves, as is typical of non-photosynthetic maize mutants.

A polyclonal antibody was raised to a recombinant fragment of ZmWHY1. This antibody detected a leaf protein whose size is consistent with that anticipated for ZmWHY1 (~25 kDa) (data not shown) and whose abundance is reduced in *ZmWhy1* mutants (Figure 1C), indicating that the detected protein is ZmWHY1. The ZmWHY1 antibody coimmunoprecipitated CRS1 (Figure 1D) from chloroplast extract, confirming that CRS1 and ZmWHY1 associate with one another. This association was disrupted by treatment with ribonuclease A (Figure 1D), indicating it is mediated by RNA. Results described below show that *atpF* intron RNA, which was shown previously to associate with CRS1 *in vivo* (19, 20), mediates the CRS1/ZmWHY1 interaction.

**Figure 1.** Mutant Alleles of *ZmWhy1*.

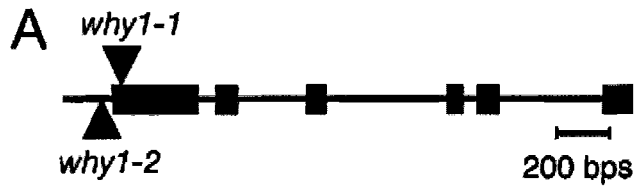
(A) Positions of *Mu* transposon insertions in the *ZmWhy1* gene. Protein coding regions are indicated by rectangles, untranslated regions and introns by lines, and *Mu* transposon insertions by triangles. The sequence of each insertion site is shown below, with the nine nucleotides that were duplicated during insertion underlined. The identity of the member of the *Mu* family is shown for each insertion (*why1-2*: Mu1/1.7; *why1-1*: MuDR), and was inferred from polymorphisms in the terminal inverted repeats.

(B) Phenotypes of *ZmWhy1* mutant seedlings grown for nine days in soil. Seedlings shown are homozygous for either the *Zmwhy1-1* or *Zmwhy1-2* allele, or are the heteroallelic progeny of a complementation cross.

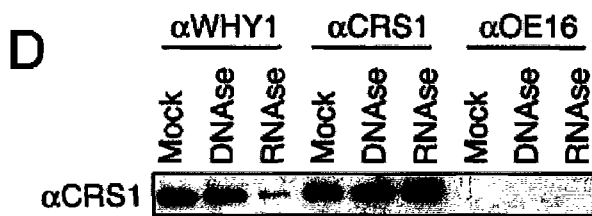
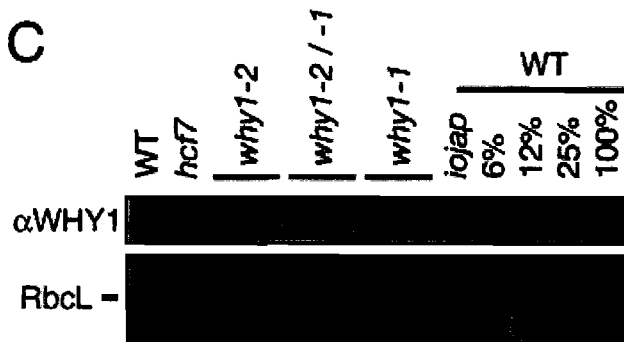
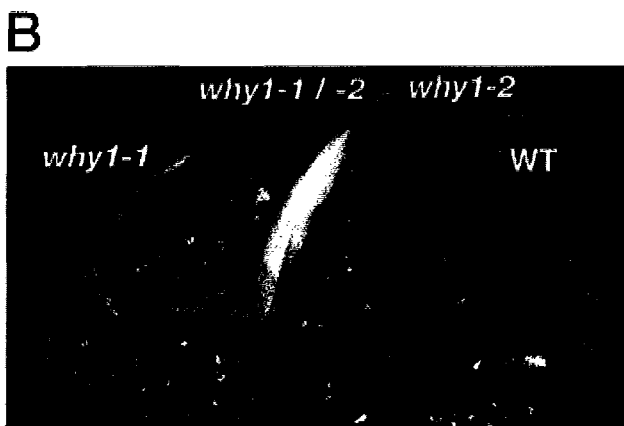
(C) Immunoblot showing loss of ZmWHY1 in mutant leaf tissue. Total leaf extract (10 µg protein, or dilutions as indicated) were analyzed. The same blot stained with Ponceau S is shown below, with the large subunit of Rubisco (RbcL) marked. *hcf7* and *iojap* are pale green and albino maize mutants with weak and severe plastid ribosome deficiencies, respectively (34, 35). The apparently higher levels of ZmWHY1 in *Zmwhy1-1* mutants relative to *Zmwhy1-2* mutants may be an artifact of the fact that samples were loaded on the basis of equal total protein: the abundant photosynthetic enzyme complexes make up the bulk of the protein in the *Zmwhy1-2* extract but are missing in the *Zmwhy1-1* extract, causing other proteins to appear over-represented.

(D) RNA dependent coimmunoprecipitation of ZmWHY1 with CRS1. Prior to immunoprecipitation, stroma was treated with DNase or RNase, or incubated under similar conditions without added nuclease (Mock). The stroma was then subjected to immunoprecipitation with the antibody named at top. Presence of CRS1 in the immunoprecipitation pellets was tested by immunoblot analysis with CRS1 antibody.





*why1-2*, *Mu 1 or 1.7* -20  
 gctgagcc gctgtctctctctcgttctcagccggtcggcgca  
 +1  
 ATGccaccgcccggcgccgctcttctctcgcctcgc ctccaagcc  
*why1-1*, *MuDR*



### **ZmWHY1 partitions between the chloroplast stroma and thylakoid membrane, to which it is bound in a DNA-dependent manner**

ZmWHY1 was initially recovered from chloroplast stroma and is predicted to localize to chloroplasts by both the TargetP (44) and Predotar (45) algorithms. Immunoblot analysis of proteins from leaf, chloroplasts, and mitochondria confirmed that ZmWHY1 is found in chloroplasts and that it is absent, or found at only very low levels, in mitochondria (Figure 2A). Analysis of chloroplast subfractions showed that ZmWHY1 is recovered in both the stromal and thylakoid membrane fractions (Figure 2A); this behavior differs from that of other chloroplast gene expression factors using the same fractionated chloroplast preparation (PPR2, PPR4, RNC1, CAF1, CAF2, CFM2), all of which were found solely in the stromal fraction (17, 19, 21, 24, 33).

It seemed possible that ZmWHY1 associated with the thylakoid membrane via a DNA tether because chloroplast nucleoids are membrane-associated (46) and AtWHY1 copurified with a chloroplast chromosome preparation (31). In support of this possibility, treatment of the thylakoid membrane fraction with DNase released a portion of the membrane-associated ZmWHY1 to the soluble fraction (Figure 2B), whereas RNase treatment had no effect. These results indicate that ZmWHY1 is associated with the thylakoid membrane, at least in part, via an association with chloroplast DNA.

### **ZmWHY1 is associated with large RNA- and DNA-containing particles**

The observations that RNase and DNase disrupt ZmWHY1's association with CRS1 and the thylakoid membrane, respectively, suggested that ZmWHY1 associates with both RNA and DNA. To further explore the nature of these interactions, the effects of RNase or DNase treatment on the sedimentation properties of ZmWHY1 were investigated (Figure 3). When untreated stroma was sedimented through a sucrose gradient, ZmWHY1 was detected in two peaks (~400-500 kDa and ~600-700 kDa) and was also found in pelleted material at the bottom of the gradient. The 600-700 kDa peak coincides with the peak of CRS1 in the same gradient. Treatment of stroma with DNase reduced the amount of ZmWHY1 in the pellet and in the ~400-500 kDa peak, but did not

reduce its recovery in the 600-700 kDa peak. Conversely, RNase treatment specifically reduced the recovery of ZmWHY1 in the 600-700 kDa peak. These results together with

**Figure 2. Intracellular Localization of ZmWHY1.**

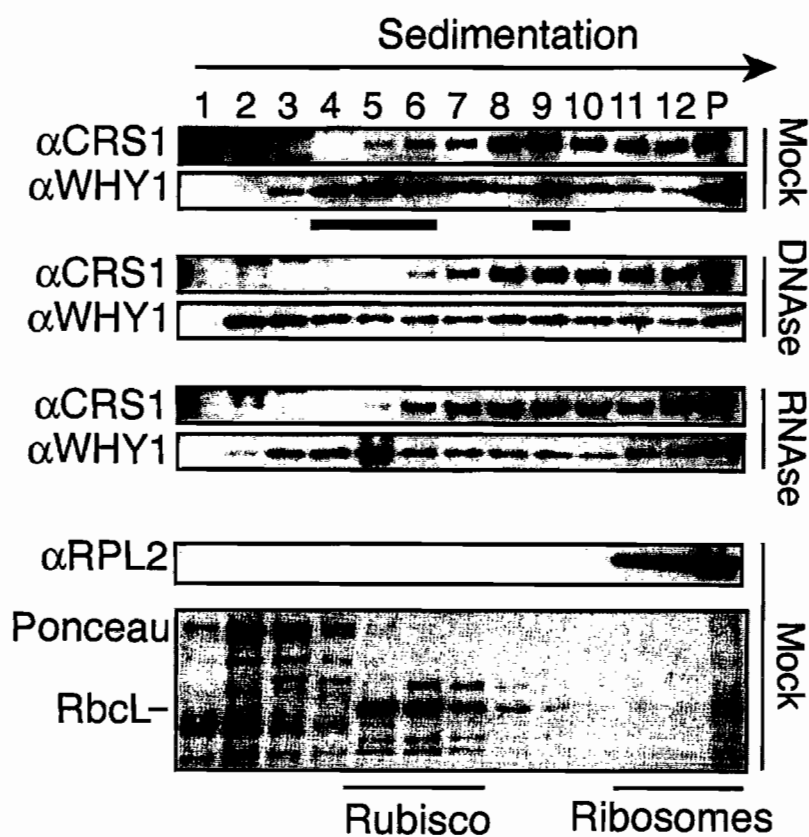
**(A)** Immunoblots of extracts from leaf and subcellular fractions. The samples in the chloroplast (Cp) and chloroplast subfraction lanes are derived from the same initial number of chloroplasts. The same blot was probed to detect a marker for thylakoid membranes (D1) and mitochondria (MDH). These subcellular fractions are the same as those shown previously for localization of RNC1, where a marker for the envelope membrane fraction was also presented (21). Env, envelope; Mito, mitochondria; Thy, thylakoid membranes. The blot stained with Ponceau S is shown below, with the band corresponding to RbcL marked.

**(B)** DNA-dependent association of ZmWHY1 with thylakoid membranes. The thylakoid membrane fraction was treated with DNase, RNase, or incubated under similar conditions without added nuclease (Mock). Thylakoid membranes were then pelleted by centrifugation. Pellet (Pel) and supernatant (Sup) fractions were brought to equal volumes, and an equivalent proportion of each fraction was analyzed on an immunoblot probed with ZmWHY1 antibody. The same blot stained with Ponceau S is shown below.



those described above suggested that ZmWHY1 resides in two types of complexes: one that includes CRS1 and RNA, and the other that includes DNA.

**Figure 3.** Sucrose-gradient sedimentation demonstrating that ZmWHY1 is associated with DNA- and RNA-containing particles in chloroplast stroma. Stromal extract was treated with DNase or RNase, or incubated under similar conditions without nuclease (Mock), and then sedimented through a sucrose gradient. An equal volume of each gradient fraction was analyzed by probing immunoblots with the antibodies indicated to the left. RPL2, a protein in the large ribosomal subunit, marks the position of ribosomes. Shown below is the blot of the mock-treated fractions stained with Ponceau S, with the RbcL band marked to illustrate the position of Rubisco. The Ponceau S stained blots of experiments involving the DNase and RNase treated extracts looked similar (data not shown).



### **Coimmunoprecipitation assays demonstrate that ZmWHY1 associates with a subset of plastid RNAs that includes the *atpF* intron**

The RNA-dependent association between ZmWHY1 and CRS1 suggested that ZmWHY might associate with CRS1's RNA ligand, the *atpF* intron. However, the albino phenotype conditioned by the *Zmwhy1-1* allele indicated that this could not be ZmWHY1's sole ligand, because mutations in *crs1* that completely block *atpF* intron splicing result in a much less severe chlorophyll deficiency (20). To identify RNAs that associate with ZmWHY *in vivo* we used a "RIP-Chip" assay (47) as an initial screen: RNAs that coimmunoprecipitate with ZmWHY1 from stromal extract were identified by hybridization to a tiling microarray of the maize chloroplast genome. To ensure that DNA associated with ZmWHY did not contribute to the signal, the extract was treated with DNase prior to immunoprecipitation, and the nucleic acids recovered from the immunoprecipitation pellet and supernatant were again treated with DNase. RNAs recovered from the pellet and supernatant were then labeled with red- or green-fluorescing dye, respectively, combined, and hybridized to the microarray. Two replicate immunoprecipitations were analyzed in this manner. To highlight sequences that are enriched in the ZmWHY1 immunoprecipitations, the median enrichment ratio [red(F635)/green (F532)] was plotted according to chromosomal position, after subtracting the median enrichment ratios from control assays (Figure 4A). The results highlight the *atpF* intron as the major RNA ligand of ZmWHY. The results suggested, in addition, an association between ZmWHY1 and RNAs derived from several other loci (e.g. *rps14*, *rpoC*, *ycf3*, *rps12*, *petD*, *rpl16*, *orf99*). When the same data were analyzed by considering only the signal in the immunoprecipitation pellets, the results were similar (Supplementary Figure 2A).

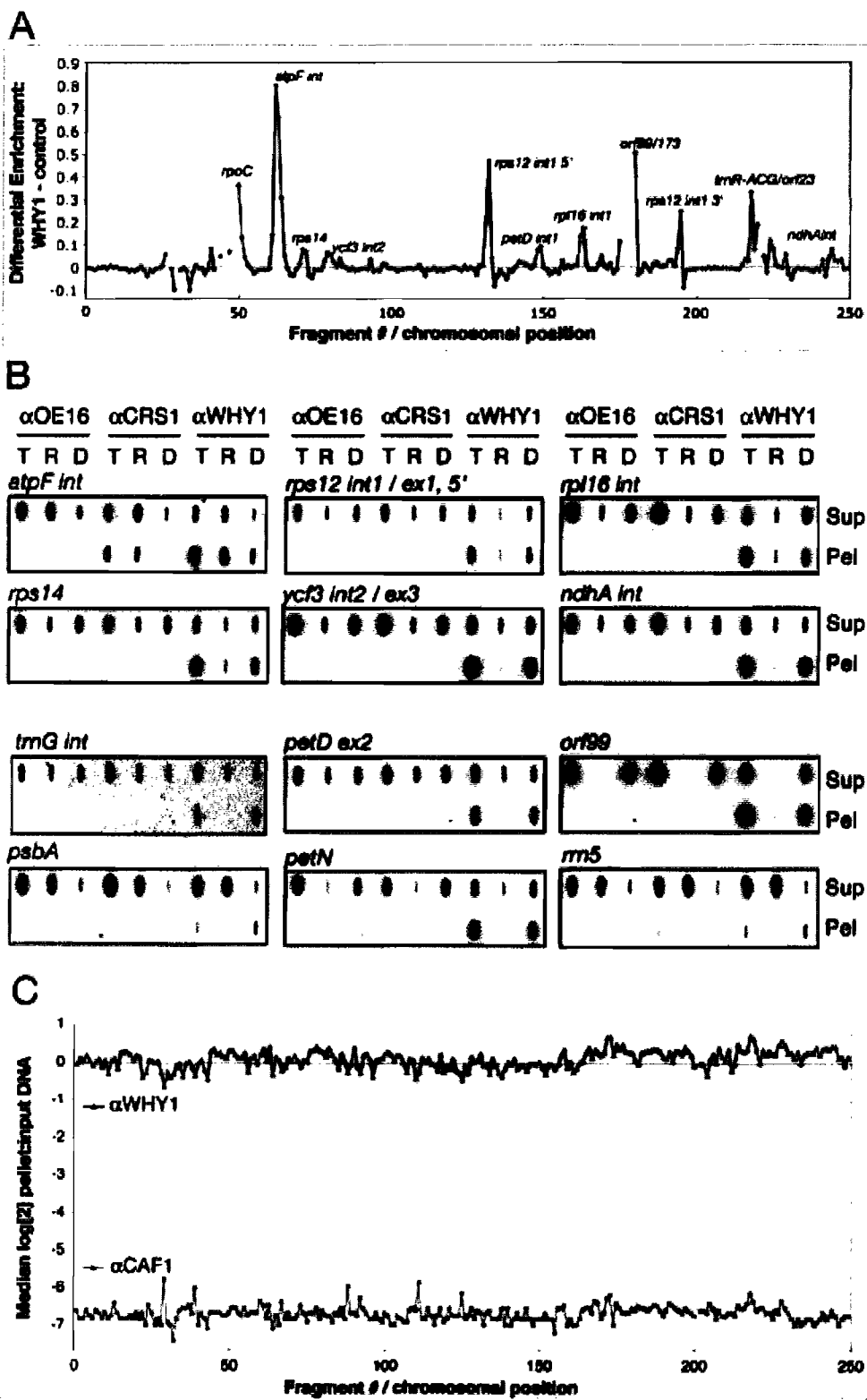
To validate candidate RNA ligands to emerge from the RIP-chip experiment, RNAs that coimmunoprecipitate with ZmWHY1 were analyzed by slot-blot hybridization using probes corresponding to each RIP-chip peak (Figure 4B). RNAs purified from immunoprecipitations with antibodies to CRS1 and OE16 (a protein that does not bind RNA) were analyzed as controls. As for the RIP-chip assays, the stromal extract was

**Figure 4.** Identification of chloroplast RNAs and DNAs that coimmunoprecipitate with ZmWHY1.

(A) RIP-chip data showing coimmunoprecipitation of specific chloroplast RNAs with ZmWHY1. The ratio of signal in the pellet versus the supernatant (F635/F532) for each array fragment is plotted according to chromosomal position. The plot shows the median values for replicate spots across two replicate ZmWHY1 immunoprecipitations after subtracting the corresponding values for two negative control immunoprecipitations (one with OE16 antibody and one without antibody). The same data are plotted using an alternative analysis method in Supplemental Figure 2B; the *atpF* intron is the most prominent peak in both analyses, but the proportional sizes of other peaks vary depending on the comparison used.

(B) Validation of RIP-chip and DIP-chip data by slot-blot hybridization. Stroma was pretreated with DNase or RNase or left untreated and then subjected to immunoprecipitation with the antibodies indicated at the top. Nucleic acids purified from the pellets (Pel) and supernatants (Sup) were further treated with DNase or alkali to remove residual DNA or RNA. The resulting total nucleic acids (T), RNA (R), or DNA (D), were applied to a nylon membrane with a slot blot manifold and hybridized with probes specific for the indicated sequences. 1/9<sup>th</sup> and 1/27<sup>th</sup> of the nucleic acid recovered from each pellet and supernatant, respectively, was applied to each slot.

(C) DIP-chip data showing genome-wide enrichment of chloroplast DNA in ZmWHY1 immunoprecipitations. Stroma was treated with RNase prior to immunoprecipitation. Nucleic acids were extracted from the immunoprecipitation pellets and from total input stroma, and subjected to alkali hydrolysis to remove residual RNA prior to analysis by microarray hybridization. The median log<sub>2</sub>-transformed ratio of fluorescence in the pellet



treated with DNase prior to immunoprecipitation and the nucleic acids recovered from the immunoprecipitation were treated again with DNase. The results largely recapitulated the RIP-chip data (see lanes “R” in Figure 4B): *atpF* intron RNA was confirmed to be strongly enriched in ZmWHY1 immunoprecipitations, whereas RNAs from the *psbA* and *petN* loci, which did not appear as positives in RIP-chip assays, likewise scored negative in the slot-blot hybridization assay. Coimmunoprecipitation with ZmWHY1 was also confirmed for RNAs from the *rps12*, *ndhA*, *rpl16*, *ycf3*, and *rps14* loci; as predicted by the RIP-chip data, their degree of enrichment was less than that for the *atpF* intron. However, RNAs from the *petD*, *orf99*, and *rrn5* loci, which appeared as minor peaks in the RIP-chip data, did not appear to be enriched based on the slot-blot data; the *orf99* transcript is of very low abundance, however, so it may be enriched in the pellet at levels that are too low to detect. These issues notwithstanding, the RIP-chip and slot-blot hybridization data together show that ZmWHY1 associates with a subset of RNAs in chloroplast extract, and that the *atpF* intron is its major RNA ligand.

#### **DNA from throughout the plastid genome coimmunoprecipitates with ZmWHY1**

The effects of DNase-treatment on ZmWHY1’s association with the thylakoid membrane (Figure 2B) and on its sedimentation rate (Figure 3) indicated that ZmWHY1 is associated with chloroplast DNA *in vivo*. To gain insight into which DNA sequences were involved in these interactions, we modified the RIP-chip protocol to detect coimmunoprecipitating DNA (DIP-chip): stromal extract was treated with ribonuclease prior to the immunoprecipitation, and alkali hydrolysis was used to remove residual RNA after the immunoprecipitation. A control immunoprecipitation used antibody to CAF1, a splicing factor that associates with specific chloroplast intron RNAs *in vivo* (19). Both ZmWHY1 and CAF1 were efficiently immunoprecipitated (Supplementary Figure 2C), but the DIP-chip data were strikingly different (Figure 4C): nearly all of the DNA in the input stromal sample coimmunoprecipitated with ZmWHY1, whereas very little DNA was recovered in CAF1 immunoprecipitations. These results confirm that ZmWHY1 is associated with chloroplast DNA and show further that ZmWHY1 either binds



throughout the chloroplast genome, or binds to specific DNA regions and coimmunoprecipitates all other DNA sequences due to their linkage to ZmWHY1 binding sites. Incubation of the extract with various restriction enzymes prior to the immunoprecipitation did not reveal the specific enrichment of any DNA sequences (Supplementary Figure 2B), leading us to favor the interpretation that ZmWHY1 is associated with many sites throughout the chloroplast genome. Nucleic acids recovered from the CAF1 and ZmWHY1 immunoprecipitations were also used as a direct template for PCR (Supplementary Figure 2D). The results support the DIP-Chip data: PCR product was obtained using a variety of chloroplast genome primers from the ZmWHY1 coimmunoprecipitation and not from the CAF1 coimmunoprecipitation.

The enrichment of DNA sequences in ZmWHY1 immunoprecipitations was further confirmed by slot blot hybridization (Figure 4B). As for the DIP-chip assays, stroma was treated with RNase prior to the immunoprecipitation, and residual RNA was removed by alkali hydrolysis after the immunoprecipitation (Figure 4B, lanes “D”). Antibody to ZmWHY1 coimmunoprecipitated DNA from all sequences tested, whereas DNA was not detected in either the CRS1 or OE16 immunoprecipitations. The DIP-chip, PCR, and slot-blot hybridization data provide strong evidence that ZmWHY1 is associated with chloroplast DNA *in vivo* and that it has many binding sites throughout the genome.

### **Zm *Why1* mutants are deficient for plastid ribosomes**

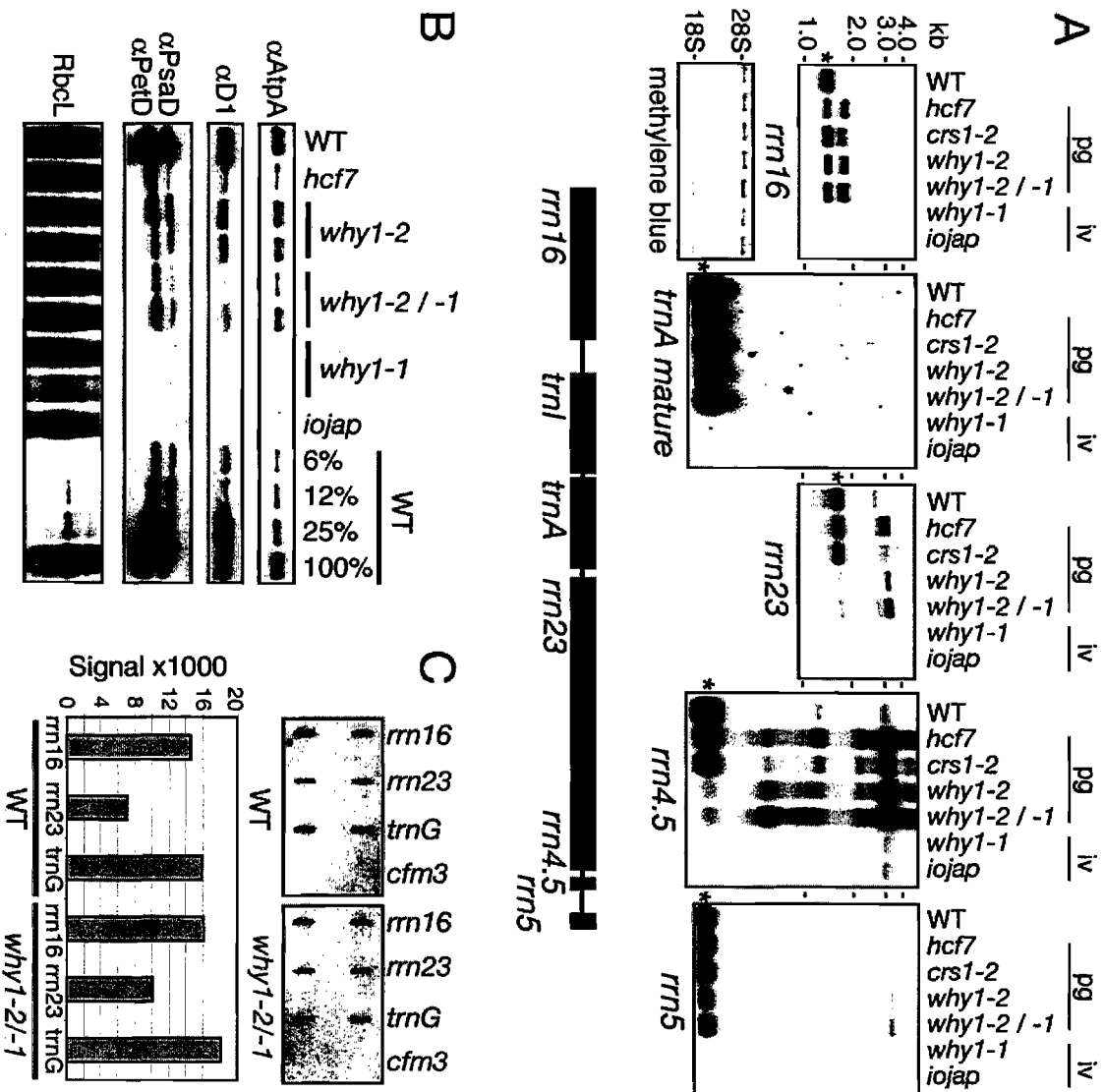
A role for WHY1 in chloroplast gene expression was suggested by the coimmunoprecipitation of ZmWHY1 with CRS1, RNA and DNA, and by the copurification of AtWHY1 with the plastid transcriptionally-active-chromosome (31). In support of this possibility, core subunits of the chloroplast ATP synthase, photosystem II, photosystem I, the cytochrome *b<sub>6</sub>f* complex, and Rubisco accumulate to reduced levels in Zm*Why1* mutants (Figure 5B). The protein deficiencies conditioned by the weak allele combinations (Zm*why1-2/-2* and Zm*why1-1/-2*) resemble those in *hcf7* mutants, which have a reduced content of chloroplast ribosomes (35).

**Figure 5.** Plastid ribosome deficiency in *ZmWhy1* mutants.

(A) Total seedling leaf RNA (0.5  $\mu$ g) was analyzed by RNA gel blot hybridization using probes for the RNAs indicated at the bottom. A map of the plastid rRNA operon is shown below. A cDNA probe was used to detect mature trnA; this lacks intron sequences and therefore hybridizes poorly to unspliced precursor. The probe for 23S rRNA is derived from the 5' portion of the *rrn23* gene and detects just one of the two 23S rRNA fragments found in ribosomes *in vivo*. The leaf pigmentation conditioned by each mutant allele is indicated: iv, ivory leaves; pg, pale green leaves. The blot used to detect 16S rRNA is shown after staining with methylene blue to illustrate equal loading of cytosolic rRNAs (18S, 28S). Mature RNA forms are indicated with asterisks.

(B) Reduced accumulation of photosynthetic enzyme complexes in *ZmWhy1* mutants. Immunoblots of leaf extract (5  $\mu$ g protein or the indicated dilutions) were probed with antibodies to core subunits of photosynthetic enzyme complexes: AtpA (ATP synthase), D1 (photosystem II), PsaD (photosystem I), and PetD (cytochrome *b<sub>6</sub>f* complex). The same blot stained with Ponceau S is shown below to illustrate sample loading and the abundance of RbcL.

(C) Plastid run-on transcription. Chloroplasts prepared from *Zmwhy1-1/-2* heteroallelic mutants or their normal siblings (WT) were used for run-on transcription assays as described in Methods. RNAs purified from the reactions were hybridized to slot blots harboring oligonucleotides corresponding to the genes indicated at the top. Each probe was present in duplicate. *cfm3*, a nuclear gene, served as a negative control. The results were quantified with a phosphorimager and plotted on the bar graph below. versus the input is plotted for replicate array fragments as a function of chromosomal position. The left inset shows the recovery of CAF1 and ZmWHY1 in the immunoprecipitations: the antibody used for immunoprecipitation is indicated above, and the antibody used to probe the immunoblot is shown to the left. Coimmunoprecipitated DNAs were also used as template for PCR using primers at several positions in the chloroplast chromosome (right inset) The fragment #s correspond to those on the microarray.



These proteins were not detectable in *Zmwhy1-1* homozygotes, as in albino *iojap* mutants which lack plastid ribosomes (Figure 5B).

The global loss of photosynthetic enzyme complexes in *ZmWhy1* mutants suggested an underlying loss of plastid ribosomes. This possibility was confirmed by RNA gel blot hybridizations, which showed a loss of mature 23S, 4.5S, and 16S rRNAs in hypomorphic *ZmWhy1* mutants, and an increased accumulation of rRNA precursors (Figure 5A). Chloroplast rRNAs were not detectable in plants homozygous for the null *Zmwhy1-1* allele, as in albino *iojap* leaves. Whereas *hcf7* mutants show a more severe loss of 16S rRNA than 23S and 4.5S rRNAs, the reverse is true for hypomorphic *ZmWhy1* mutants. A dramatic increase in the ratio of 23S rRNA precursors to mature 23S rRNA in these mutants was confirmed with a poisoned-primer extension assay (Supplementary Figure 3C).

Some steps in rRNA processing are dependent upon ribosome assembly in chloroplasts, as in bacteria (see, for example, refs. (24, 35, 48)). The aberrant 23S and 4.5S rRNA processing in *ZmWhy1* mutants suggested therefore that ZmWHY1 might promote the expression of a gene needed for the assembly of the large ribosomal subunit (an rRNA or ribosomal protein), with loss of the small ribosomal subunit being a secondary effect. It seemed plausible, for example, that ZmWHY1 might promote processive transcription through the chloroplast *rrn* operon; this would differentially affect the large ribosomal subunit due to the distal position of the genes encoding its rRNA components (23S, 4.5S, and 5S rRNA) in the operon (see map in Figure 5A). However, the results of chloroplast run-on transcriptions assays argue against this possibility (Figure 5C): the ratio of polymerase transit through the 23S gene in comparison to the 16S rRNA gene, and the ratio of *rrn* operon transcription in comparison to transcription from a different chloroplast locus (*trnG-UCC*) were similar in wild-type and *Zmwhy1-1/-2* mutant chloroplasts. Furthermore, the rRNA components of the large ribosomal subunit were not reproducibly enriched in ZmWHY co-immunoprecipitates (Figure 4A, Supplementary Figure 2B); this suggests that ZmWHY1 does not interact directly with rRNAs or 50S ribosomal subunits, although such

interactions cannot be eliminated based on these negative results. Taken together, these results argue that ZmWHY1 directly impacts the expression of a gene encoding a component of the large ribosomal subunit and/or promotes ribosome assembly. Elucidation of its precise role in this process will require further study.

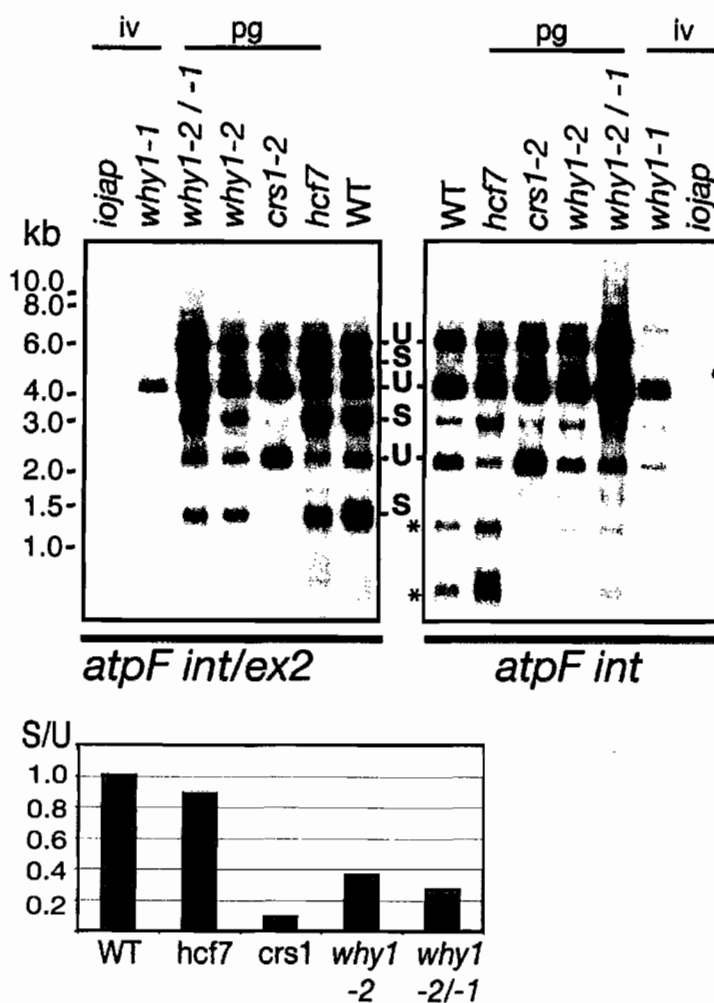
### **ZmWHY1 promotes *atpF* intron splicing**

The coimmunoprecipitation of ZmWHY1 with the *atpF* splicing factor CRS1 and with RNA from the *atpF* locus suggested that ZmWHY1 might be involved in the splicing of *atpF* pre-mRNA. To test this possibility, *atpF* RNA from *Zmwhy1* mutants was analyzed by RNA gel blot hybridization (Figure 6). To control for pleiotropic effects of weak and severe plastid ribosome deficiencies, RNAs in pale green (hypomorphic) *Zmwhy1-2* and *Zmwhy1-2/-1* mutants were compared to those in *hcf7* mutants, and RNAs in albino (null) *Zmwhy1-1* mutants were compared to those in *iojap* mutants. These comparisons were important because the complete absence of plastid ribosomes results in the failure to splice all chloroplast subgroup IIA introns, including the *atpF* intron (15, 49, 50).

The results in Figure 6 show that the ratio of spliced (S) to unspliced (U) *atpF* transcripts is reduced in hypomorphic *ZmWhy1* mutants in comparison to wild-type and *hcf7* plants, albeit not as severely as in *crs1* mutants. The ratio of excised intron (asterisks) to unspliced RNA is also reduced, supporting the interpretation that ZmWHY1 promotes *atpF* splicing rather than stabilizing the spliced product. The normal splicing of the *atpF* intron in *hcf7* mutants argues that the partial plastid ribosome deficiency in hypomorphic *ZmWhy1* mutants cannot account for their reduced *atpF* splicing. Furthermore, a different subgroup IIA intron, the *rpl2* intron, is spliced normally in the same plants (Supplementary Figure 3B), showing that not all subgroup IIA introns are affected in the hypomorphic *ZmWhy1* mutants. These results provide strong evidence that ZmWHY1's association with *atpF* RNA enhances the splicing of the *atpF* intron.

**Figure 6.** Reduced *atpF* intron splicing in *ZmWhy1* mutants.

RNA gel blot analysis of *atpF* splicing. Total seedling leaf RNA (5  $\mu$ g) was analyzed by RNA gel blot analysis using a probe including *atpF* exon 2 and a portion of the *atpF* intron (*atpF int/ex2*), or with an intron-specific probe (*atpF int*). The *atpF* gene is part of a polycistronic transcription unit that gives rise to a previously-characterized population of RNAs (51, 52). Spliced (S) and unspliced (U) transcripts are indicated. Asterisks mark bands that we believe correspond to the excised intron and its degradation products. The ratio of spliced to unspliced transcripts was quantified with a phosphorimager, normalized to the wild-type ratio, and plotted below using arbitrary units.



The coimmunoprecipitation data demonstrated an association between ZmWHY1 and RNAs from several loci other than *atpF*. However, RNA gel blot hybridizations showed that the transcripts from all such genes were qualitatively similar in *ZmWhy1* mutants and in the relevant control mutant (Figure 7). The coimmunoprecipitation of ZmWHY1 with RNAs from both loci encoding the trans-spliced group II intron in *rps12* was intriguing (Figure 4A), but splicing of this RNA is not disrupted in *ZmWhy1* mutants (Supplementary Figure 3B). These results show that ZmWHY1 is not necessary for the normal processing of most chloroplast transcripts.

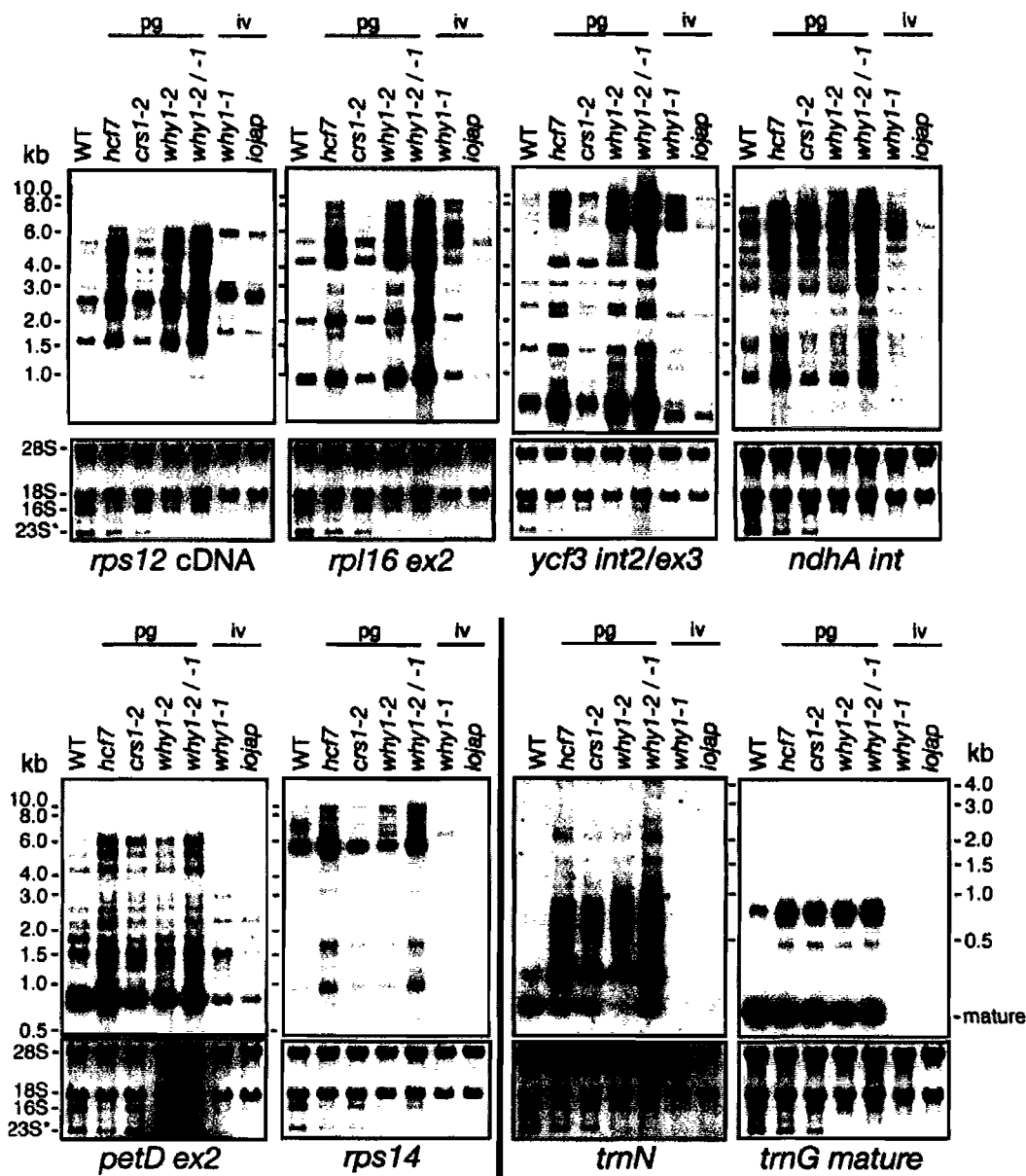
A structural homolog of ZmWHY1 in *Trypanosoma brucei* is required for mitochondrial RNA editing (53). Several plastid RNAs that are known to be substrates for RNA editing were represented among the RNAs that coimmunoprecipitate with ZmWHY1. Direct sequencing of RT-PCR products demonstrated, however, that the editing of the known edited sites in the *petB*, *rpl20*, *ycf3*, and *rps14* transcripts is not disrupted in *Zmwhy1-1* and *Zmwhy1-2/-1* mutants (data not shown), suggesting that ZmWHY1 is not required for RNA editing.

### **ZmWHY1 is required neither for chloroplast DNA replication nor for global plastid transcription**

The association of ZmWHY1 with plastid DNA suggested that it might be involved in chloroplast transcription or DNA replication. However, Southern blot analysis of total leaf DNA showed that plastid DNA levels in *ZmWhy1* mutants, although somewhat variable from sample to sample, were generally similar to those in normal and control mutant plants (Figure 8). In addition to the plastid transcripts shown in Figure 7, a variety of other transcripts were examined by RNA gel blot hybridization (Supplementary Figure 3A). In no case was a significant reduction in transcript level detected, indicating that ZmWHY1 is not necessary for global plastid transcription. In fact, a trend is apparent toward increased transcript abundance in *ZmWhy1* mutants, but these changes are rather subtle and indirect effects on RNA abundance cannot be excluded.

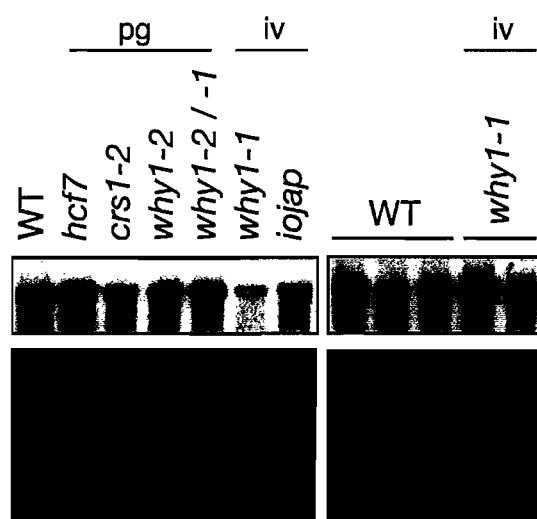
**Figure 7.** Accumulation of plastid RNAs in *ZmWhy1* mutants.

Total seedling leaf RNA (5  $\mu$ g) was analyzed by RNA gel blot hybridization using probes specific for the RNAs indicated at bottom. The *rps12* probe was a cDNA probe containing exons 1 and 2. The leaf pigmentation conditioned by each mutant allele is indicated: iv, ivory; pg, pale green. The methylene blue-stained blots are shown below, with rRNAs marked. Additional RNAs that were analyzed analogously are shown in Supplementary Figure 3A.





**Figure 8.** Chloroplast DNA levels in *ZmWhy1* mutants. Seedling leaf DNA (5  $\mu$ g) was digested with EcoRI (left), or PvuII (right) and analyzed by DNA gel blot hybridization using a probe from the chloroplast *rrn23* gene (top left), or orf99 (top right). The same gels stained with ethidium bromide are shown below. The small fluctuations in relative band intensity may result from small differences in sample loading.



**Recombinant ZmWHY1 binds single-stranded RNA and DNA *in vitro***

To determine whether ZmWHY1 can directly bind both RNA and DNA, recombinant ZmWHY1 (rWHY1) was generated by expression as a maltose binding protein (MBP) fusion. rWHY1 was released from the MBP moiety by protease cleavage and further purified on a gel filtration column (Figure 9A). rWHY1 eluted from the sizing column at a position corresponding to a globular protein of ~100 kDa, consistent with the report that StWHY1 forms a homo-tetramer (28). Filter binding assays showed that rWHY1 binds to unspliced *atpF* RNA *in vitro* (Figure 9B), but it did not show specificity for this RNA relative to other RNAs of similar size under the conditions tested (data not shown).

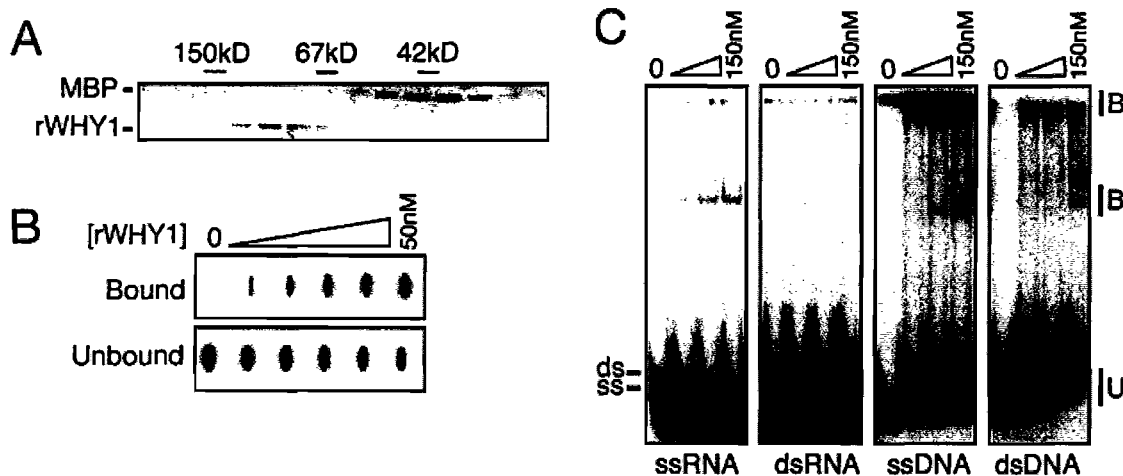
To compare the affinity of ZmWHY1 for single-stranded and double-stranded RNA and DNA, gel mobility shift assays were used to detect binding to a synthetic 31-mer oligonucleotide in the context of single-stranded DNA, single-stranded RNA, double-stranded DNA, or double-stranded RNA (Figure 9C). ZmWHY1 bound rather weakly to these short oligonucleotides but the results showed, nonetheless, that rWHY1 binds both ssDNA and ssRNA, and binds poorly to dsRNA and dsDNA.

**Figure 9.** Recombinant ZmWHY1 binds single-stranded RNA and DNA.

(A) Elution of recombinant ZmWHY1 from a gel filtration column. MBP-WHY1 was purified by amylose affinity chromatography, cleaved with TEV protease to separate the WHY1 and MBP moieties, and applied to a Superdex 200 column. Column fractions were analyzed by SDS-PAGE and staining with Coomassie blue. The elution position of size markers (alcohol dehydrogenase, 150 kDa; BSA, 67 kDa, MBP, 42 kDa) is shown. The peak WHY1 fractions were pooled and used for *in vitro* assays.

(B) Filter binding assay showing RNA binding activity of ZmWHY1. Assays containing 10 pM radiolabeled *atpF* intron RNA and increasing ZmWHY1 concentrations (50 nM maximum) were filtered through sandwiched nitrocellulose and nylon membranes. Protein/RNA complexes were captured on the nitrocellulose (bound); unbound RNA was captured on the nylon membrane below.

(C) Gel mobility shift assay showing rWHY1's relative affinity for double and single stranded RNA and DNA. A 31-mer oligonucleotide in RNA or DNA form was radiolabeled, heated, and either snap cooled (ssRNA, ssDNA), or cooled slowly in the presence of monovalent salts and a two fold excess of its complement (dsRNA, dsDNA). The substrate (40 pM) was mixed with increasing concentrations of ZmWHY1 (17, 50, 150 nM). Protein binding is illustrated by the appearance of an upper band and retention at the top of the gel, and by the disappearance of unbound substrate.



## Discussion

Previous reports have attributed diverse functions and intracellular locations to WHY1. WHY1 in dicots has been reported to be a single-stranded DNA binding protein that functions in the nucleus as both a transcription factor (26, 28) and as a negative regulator of telomere length (29). Arabidopsis WHY1 copurified with the “transcriptionally active chromosome” from chloroplasts (31). Our results add another layer to this complex picture. We demonstrate that ZmWHY1 is essential for chloroplast biogenesis, and that it localizes to the chloroplast where it plays multiple roles in gene expression. We also add RNA-binding to WHY1’s repertoire of biochemical activities and demonstrate that ZmWHY1 is bound to a subset of chloroplast RNAs in chloroplast extract.

### Multiple roles of ZmWHY1 in chloroplast biogenesis

ZmWHY1 was identified among proteins that coimmunoprecipitate with CRS1, which is required for the splicing of the group II intron in the chloroplast *atpF* pre-mRNA. We showed that ZmWHY1 is associated with *atpF* intron RNA *in vivo* and that the coimmunoprecipitation of ZmWHY1 and CRS1 is disrupted by RNase, indicating that they coimmunoprecipitate due to their association with the same RNA molecule. ZmWHY1’s association with *atpF* RNA is functionally significant, as *atpF* intron splicing is disrupted in *ZmWhy1* mutants. However, the splicing of this intron is more sensitive to a partial loss of CRS1 than to a partial loss of ZmWHY1, suggesting that ZmWHY1 plays an accessory function in *atpF* splicing but may not be absolutely required.

The *atpF* splicing defect in *Zm Why1* mutants cannot account for their loss of plastid ribosomes, as the more severe *atpF* splicing defect in *crs1-1* mutants is not accompanied by a substantial plastid ribosome deficiency (20). The specific role of ZmWHY1 in promoting the biogenesis of the plastid translation machinery remains unclear. Although several RNAs with translation-related functions are among the RNAs

that coimmunoprecipitate with ZmWHY1, the abundance and processing of these RNAs are similar in *ZmWhy1* mutants and in control mutants that exhibit a ribosome-deficiency of similar severity. The specific rRNA deficiencies in *ZmWhy1* mutants do suggest, however, that ZmWHY1 is most directly involved in the biogenesis of the large ribosomal subunit: the accumulation and processing of the 23S and 4.5S rRNAs are more sensitive to the partial loss of *ZmWhy1* function than are those of 16S rRNA, whereas the reverse is true for *hcf7* mutants. Furthermore, in *ppr5* mutants, whose primary defect is in the maturation of a specific plastid tRNA, the rRNAs from the two ribosomal subunits are impacted to a similar extent (48). Thus, our results point to the biogenesis of the plastid large ribosomal subunit as one function of ZmWHY1 but definition of its precise role in this process will require additional study. The strong defect in the processing step that separates 23S rRNA from 4.5S rRNA in hypomorphic *ZmWhy1* mutants is reminiscent of defects reported for mutations in the *DCL*, *DAL*, and *RNR1* genes in dicots (54-57). Although it is unclear whether any of these genes functions directly in 23S/4.5S rRNA processing, it is possible that WHY1 acts in concert with one or more of these proteins.

### **ZmWHY1 binds both RNA and DNA *in vitro* and *in vivo***

We show here that chloroplast DNA coimmunoprecipitates with ZmWHY1 from plastid extract, that a fraction of ZmWHY1 is tethered to the thylakoid membrane in a DNA-dependent fashion, that a fraction of stromal ZmWHY1 is found in DNA-containing particles of ~400 kDa, and that Zm WHY1 binds single-stranded DNA *in vitro*. These results are consistent with previous reports that dicot WHY1 binds single-stranded DNA (28, 29) and that it copurifies with a chloroplast “transcriptionally-active chromosome” (31). Our findings suggest that ZmWHY1 either binds DNA in a sequence non-specific fashion or that it has many binding sites distributed throughout the plastid genome, because DNA sequences from throughout the plastid genome coimmunoprecipitated to a similar extent with ZmWHY1. It remains possible, however, that ZmWHY1 associates with specific DNA regions *in vivo*, but that these associations

were disrupted during lysate preparation. A DNA-immunoprecipitation experiment was recently reported for AtWHY2, a mitochondrial-localized Whirly protein (32), with analogous results: DNA sequences from a variety of regions throughout the mitochondrial genome coimmunoprecipitated with AtWHY2, when assayed by PCR.

We demonstrate here that ZmWHY1 interacts not only with DNA, as anticipated by previous reports, but that it also binds RNA *in vivo* and *in vitro*. That ZmWHY1 interacts with RNA is, perhaps, not surprising given that a structural homolog of ZmWHY1 has been shown to bind RNAs involved in kinetoplastid RNA editing (53), and that many proteins that bind single-stranded DNA also bind RNA. The *atpF* intron RNA was the major RNA ligand of ZmWHY1 detected in the RNA coimmunoprecipitation assays. This RNA is not particularly abundant *in vivo* so its enrichment in ZmWHY1 immunoprecipitations likely reflects a specific interaction *in vivo*. Although intrinsic specificity for this RNA did not emerge from *in vitro* binding assays using the entire intron, a high affinity site within a large RNA such as the *atpF* intron (~800 nucleotides) can be masked *in vitro* due to the over-whelming number of non-specific sites available for protein binding. Therefore, more detailed studies involving smaller RNA ligands will be required to determine whether ZmWHY1 binds RNA with sequence-specificity, or whether it is recruited to the *atpF* intron via protein-protein interactions.

### **What is WHY1's DNA-related function in the chloroplast?**

The association of ZmyWHY1 with DNA sequences from throughout the chloroplast genome suggests that it participates in transcription and/or DNA metabolism. However, our results argue against a general role in transcription, as all plastid mRNAs examined accumulate in hypomorphic *Zmwhy1* mutants to levels that are comparable to those in the relevant control mutants. The results of chloroplast transcription run-on experiments argue that the preferential loss of 23S rRNA in these mutants is due to aberrant ribosome assembly rather than to reduced rRNA transcription rates. It remains

possible, however, that ZmWHY1 does play a role in chloroplast transcription but that another gene with a partially redundant function serves this purpose in *ZmWhy1* mutants.

It is intriguing that ZmWHY1 binds preferentially to DNA in single-stranded form because opportunities to interact with single-stranded DNA *in vivo* are expected to be limited. DNA replication, recombination and repair involve the transient occurrence of single-stranded DNA, and torsional stress can induce DNA unwinding. The Southern blot data showing that plastid DNA levels are no more than minimally decreased in *ZmWhy1* null mutants argue against a central role for ZmWHY1 in DNA replication; however participation of ZmWHY1 in DNA recombination or repair remains possible. In fact, the participation of an unrelated ssDNA binding protein, OSB1, in plant mitochondrial DNA recombination was reported recently (58).

There are several parallels between our findings with ZmWHY1 and the activities reported for the bacterial protein HU. HU is associated with the bacterial nucleoid, binds preferentially to DNA with irregular structural features (e.g. single-stranded gaps and bulges), and is involved in DNA recombination and repair (59, 60). Despite its high conservation in bacteria and the presence of an HU homolog in a plastid genome in red algae (61), HU homologs are not encoded in the nuclear or plastid genomes of vascular plants (61, 62). Thus, alternative proteins have presumably been recruited in vascular plants to fulfill the functions performed by HU in the chloroplast's cyanobacterial ancestor. The nucleoid-associated protein sulfite reductase has been suggested to be one such protein (62-64), and perhaps WHY1 is another. HU influences global transcription patterns through its effect on nucleoid architecture, and mediates the formation of DNA loops that repress transcription from specific genes (65-67). HU is also an RNA binding protein, and functions *in vivo* to repress the translation of the *E. coli rpoS* mRNA (68, 69). Like HU, ZmWHY1 interacts globally with plastid DNA, but specifically with certain plastid RNAs, and binds preferentially to nucleic acids with single-stranded character. The abundance of several chloroplast mRNAs is increased in *ZmWhy1* mutants, consistent with a global repressive role for ZmWHY1 in transcription. This possibility is in accord with the recent report that over-expression of AtWHY2 in

*Arabidopsis* causes a reduction in the levels of several mitochondrial RNAs (32). Although its role in DNA metabolism remains uncertain, our results demonstrate that description of WHY1 as a chloroplast transcription factor is, at best, an oversimplification of the complex roles played by this interesting protein.

### Bridge

The preceding chapter discusses WHY1, a plant specific RNA and DNA binding protein in the Whirly protein family. The severe phenotype of *WHY1* mutant plants suggests this protein is crucial for chloroplast biogenesis, however the mechanism of WHY1s function is still not understood. The following chapter will discuss the pentatricopeptide repeat (PPR) family, another family of proteins important for organelle biogenesis. Like WHY1, PPR proteins are sequence-specific binders of RNA, and are indispensable for chloroplast function. Both Whirly family members and PPR family member are targeted to either the chloroplasts or the mitochondria. Whereas Whirly proteins comprise a small, plant specific family (2 to 3 members per species), PPR proteins are found in all eukaryotes, and the family is extremely large in plants, consisting of more than 450 members in angiosperms (11). The data presented here gives insight into how this diverse family of proteins may function to regulate chloroplast and mitochondrial gene expression.



## CHAPTER III

### **BIOCHEMICAL ANALYSES SUGGEST THAT PPR/RNA INTERACTIONS INVOLVE AN UNUSUAL RNA/PROTEIN INTERFACE THAT IS SUFFICIENT TO MEDIATE A VARIETY OF POSTTRANSCRIPTIONAL EFFECTS**

This chapter describes analyses of two members of the pentatricopeptide repeat protein family, PPR10 and PPR5. This work was done in collaboration with Dr. Alice Barkan, Margarita Rojas, and Omer Ali Bayraktar. Margarita Rojas performed the structure probing assays and some of the partial alkali hydrolysis binding assays, and Omer Ali Bayraktar performed the PPR10 partial alkali hydrolysis binding assay with 5' end labeled RNA.

#### **Introduction**

Mitochondria and chloroplasts contain small genomes that reflect their origins as free-living bacteria. The organellar genomes are much reduced in comparison to those in their bacterial ancestors, and their gene expression mechanisms have diverged considerably. For example, genes in chloroplasts are transcribed by two different types of RNA polymerase, and the transcripts are then subject to an array of processing events that include RNA editing, group I and group II intron splicing, and the processing of polycistronic precursors to yield monocistronic mRNAs. These events are carried out by nucleus-encoded proteins, most of which are innovations that evolved in the eukaryotic host.

The pentatricopeptide repeat (PPR) family is a notable example of a host-derived protein family that mediates gene expression in chloroplasts and mitochondria (reviewed in 70). PPR proteins consist of up to ~25 degenerate repeats of a 35 amino acid sequence, usually in a single tandem array (10). They are found in all eukaryotes but form a greatly expanded family in plants, with more than 450 members in angiosperms (11). The PPR motif shares homology with the TPR motif, a helical hairpin motif found in repeated arrays that mediates protein-protein interactions. However, genetic data have consistently implicated PPR proteins in functions related to RNA metabolism; these include RNA editing, RNA splicing, RNA cleavage, RNA stabilization, and translational control (reviewed in 70). Biochemical analyses of several PPR proteins support the notion that they exert downstream effects through site-specific binding to RNA (71-73). However, the mechanistic basis of the diverse activities attributed to PPR proteins is largely unexplored.

To elucidate how PPR proteins recognize specific RNA sequences and mediate their effects on RNA metabolism, we are studying several PPR/RNA interactions in detail. We describe here *in vitro* analyses of two chloroplast PPR proteins, PPR5 and PPR10, whose physiological functions and *in vivo* binding sites were reported previously. PPR5 binds within a group II intron found in a chloroplast tRNA precursor (trnG-UCC), protecting it from inactivation by an endonucleolytic cleavage (48, 71). PPR10, in contrast, binds in the intergenic regions of two polycistronic transcripts and stabilizes adjacent RNA segments. That PPR10 binding sites are found at the immediate 5' or 3' ends of those RNAs it stabilizes suggested that PPR10 serves as a barrier to exonucleolytic RNA degradation *in vivo* (72).

Results presented here provide evidence that PPR10 is sufficient to block RNA degradation by both 3'→5' and 5'→3' exoribonucleases *in vitro*. In addition, we define the minimal RNA segments required for a high affinity interaction with PPR5 and PPR10, and probe the effects of these interactions on adjacent RNA structures. The results support the notion that PPR5 and PPR10 bind an extended stretch of single-stranded RNA, and that this binding disrupts RNA structures that would otherwise inhibit

splicing and translation, respectively. These findings suggest plausible mechanisms underlying the ability of PPR10 to enhance *atpH* translation (72) and PPR5 to enhance trnG-UCC splicing *in vivo* (48). This study shows how two seemingly disparate functions of PPR proteins, translational activation and promotion of splicing, can be explained as a passive consequence of the ability of a PPR tract to bind in a sequence-specific fashion to an extended segment of single-stranded RNA. We speculate that most or all of the functions attributed to proteins comprised purely of PPR repeats may result from their intrinsic ability to block access to the RNA by other proteins and to remodel RNA structures.

## Materials and Methods

### Ribonucleic acid binding assays

Gel mobility shift (GMS) assays were performed as previously described (71). Briefly, *in vitro* transcribed RNAs (oligonucleotides 3,4,5,8, and 9 in the PPR5 assays) or synthetic RNAs (all PPR10 oligonucleotides and oligonucleotides 1,2,6, and 7 used for PPR5 GMS assays) were 5'-end labeled with [ $\gamma$ -<sup>32</sup>P]-ATP. PPR10 binding reactions contained 100 mM NaCl, 40 mM Tris pH 7.5, 4 mM DTT, 0.1 mg/ml BSA, 0.5 mg/ml heparin, 10% glycerol, 10 units RNAsin, ~40 pM radiolabeled RNA, and protein concentrations as indicated. The PPR10 stoichiometric binding assay was performed as for the PPR10 GMS assays, except that it included 100 nM RNA (~40 pM radiolabeled, the rest was unlabeled)(19 nt sequence shown in Figure 1) and increasing concentrations of protein as indicated. PPR5 binding reactions contained 100mM NaCl, 1 mg/ml heparin, 40 mM Tris pH 7.5, 4 mM DTT, .04 mg/ml BSA, 10% glycerol, 10 units RNAsin, ~40 pM radiolabeled RNA, and protein concentrations as indicated. All reactions were incubated for 20 min at 25°C and resolved on 5% native polyacrylamide gels.

### **Minimal binding assay using partially alkali hydrolyzed RNA**

10 pmols of 5' or 3' end label RNA oligonucleotide (55nt trnG intron RNA for PPR5 and 49 nt atpH 5' UTR RNA for PPR10) were ethanol precipitated and resuspended in alkaline hydrolysis buffer (50mM Na<sub>2</sub>CO<sub>3</sub> pH9.5 and 1mM EDTA). The RNA was distributed in 5 different tubes and boil for 1,2,3,4 and 5 min respectively, and then snap cooled on ice for 1 min. The RNA was purified by phenol: chloroform extraction, and ethanol precipitation. The hydrolyzed RNA was incubated in the absence or presence of recombinant protein at 20°C for 20 min under the following buffer conditions: 30mM Tris pH7.5, 100mM NaCl, 4mM DTT, 0.04mg/ml BSA, and 500ng/μl of heparin (25 ng/μl heparin for PPR10). Binding reactions were separated on a 5% native polyacrylamide gel in 1x THE buffer as previously described (71). The set of bands corresponding to the bound and unbound fractions were excised, eluted in RNA elution buffer (0.5M NH<sub>4</sub>OAc, 0.25% SDS, 1mM EDTA), extracted with phenol: chloroform, and precipitated with Ethanol. Samples were resuspended in 20μl formamide loading dye and analyzed on an 8% polyacrylamide gel in 1XTBE as previously described (71).

### **PNPase purification**

His tagged *Synechocystis* polynucleotide phosphorylase (PNPase) expression construct in pET-20b (+) vector was generously provided by the Shuster lab. PNPase was expressed in BL21 star *E.coli* cells. Induction and lysis via sonication were preformed as described in Williams-Carrier *et al.* (2008) except that lysis buffer consisted of 50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 8, 300 mM NaCl, 20 mM imidazole, 10% glycerol, 1% Tween-20, and 2 mM BME. The lysate was cleared by centrifugation at 13,000 g for 20 min. Cleared lysate was bound to 1 ml Ni-NTA agarose (Qiagen) and incubated for 1 h at 4°C. Slurry was put on .8X4 cm Poly-Prep Chromatograph Column (Bio-Rad). Column was washed 3 times with 5 ml of lysis buffer. Protein was eluted with 1 ml lysis buffer containing 100 mM imidazole, followed by 2 ml lysis buffer containing 250 mM imidazole. Elute was brought up to 15.5 ml volume with Q buffer (20 mM HEPES pH8, 50 mM NaCl, 12.5

mM MgCl<sub>2</sub>, .1 mM EDTA, 2 mM DTT). Elute was then added to 1 ml Q Sepharose, Fast Flow (Amersham Biosciences). Slurry was incubated for 1 h at 4°C and put on .8X4 cm Poly-Prep Chromatograph Column (Bio-Rad). Column was washed 3 times with 5 ml of Q buffer. Protein was eluted through a series of 1 ml washes with Q buffer containing increasing concentrations of NaCl: 150 mM, 300 mM, 450 mM, and 600 mM. Protein eluted at ~300mM NaCl. Buffered glycerol (~100% glycerol with Q buffer constituents) was added to a final concentration of 19%. Protein aliquots were taken and stored at -20°C for use, and -80°C for long-term storage.

### ***In vitro* exonuclease protection assays**

PNPase assays for 3'→5' exonuclease activity. Synthetic RNA oligonucleotide corresponding to the atpH 5' UTR (sequence in Figure 3A) was 5'-end labeled with [ $\gamma$ -<sup>32</sup>P]-ATP and gel purified as for the gel mobility shift assays (71). ~80 pM radiolabeled RNA was heated 2 min at 90°C, removed from heat, and snap cooled on ice. Salt mix was added to final concentration of 25  $\mu$ g/ml Heparin, 30 mM Tris pH 7.5, 100mM NaCl, 4 mM DTT. 5  $\mu$ l PPR10 was added to PPR10 + samples, final concentration 100 nM. PPR10 dialysis buffer was added to PPR10 – samples. Final sample volume was 25  $\mu$ l. Samples were incubated 15 min at 25°C. 2 $\mu$ l PNPase was added to PNPase + samples, final concentration 440 nM. PNPase buffer (Q buffer with 19% glycerol) was added to PNPase – samples. Samples were incubated at 25°C for 20 min. 10  $\mu$ l of each sample was run on a 5% native gel as in the gel mobility shift assays (71). Remaining sample was phenol extracted and ethanol precipitated. RNA pellets were resuspended in 15  $\mu$ l formamide die mix boiled 3 min and applied to a 30 cm long, 8% polyacrylamide, 8M urea, 1X TBE (89 mM Tris pH 8.3, 89 mM boric acid, 2 mM EDTA), denaturing gel. Gels were run at 20 W (constant power) at room temperature until the bromophenol blue dye migrated to ~8 cm from the bottom of the gel.

Terminator exonuclease assays for 5'→3' exonuclease activity. Synthetic RNA oligonucleotide corresponding to the atpH 5' UTR (sequence in Figure 3A) was 3'-end labeled by annealing with a DNA sequence complementary to the last (3') 20 nt and

beginning with an additional 5' G. Klenow polymerase lacking the exonuclease domain was used to incorporate an [ $\alpha$ - $^{32}$ P]-CTP. The product was gel purified as for the gel mobility shift assays (71). ~80 pM radiolabeled RNA was heated 2 min at 90°C, removed from heat, and snap cooled on ice. Salt mix was added to final concentration of 25  $\mu$ g/ml Heparin, 50 mM Tris pH 8, 100mM NaCl, 2 mM MgCl<sub>2</sub>, 4 mM DTT. 5  $\mu$ l PPR10 was added to PPR10 + samples, final concentration 100 nM. PPR10 dialysis buffer was added to PPR10 – samples. Final sample volume was 25  $\mu$ l. Samples were incubated 15 min at 25°C. 2 $\mu$ l Terminator 5'→3' exonuclease (Epicentre Biotechnologies) was added to Terminator + samples, final concentration 100 nM. Samples were incubated at 25°C for 20 min. 10  $\mu$ l of each sample was run on a 5% native gel as in the gel mobility shift assays (71). Remaining sample was phenol extracted and ethanol precipitated. RNA pellets were resuspended in 15  $\mu$ l formamide die mix boiled 3 min and applied to a 30 cm long, 8% polyacrylamide, 8M urea, 1X TBE (89 mM Tris pH 8.3, 89 mM boric acid, 2 mM EDTA), denaturing gel. Gels were run at 20 W (constant power) at room temperature until the bromophenol blue dye migrated to ~8 cm from the bottom of the gel.

### **Nuclease cleavage structure probing assays**

5-end labeled trnG 55mer RNA oligonucleotide (0.1pmols) in the absence or presence of rPPR5 protein was incubated at 20°C for 20 min under the following buffer conditions 30mM Tris pH7.5, 100mM NaCl, 4mM DTT, 0.04mg/ml BSA, and 100ng/ $\mu$ l of heparin. The binding step was followed by cleavage with varying concentrations of either RNaseT1 (Ambion) or RNase V1 (Ambion) or Mung Bean Nuclease (NEB) or Rnase H (Ambion) at 20°C for 10 min. Treated RNA was added to 10 $\mu$ l of formamide loading dye. Samples were analyzed on either and 15% or an 8% polyacrylamide, 8M urea gels. A limited alkaline digestion of trnG55 mer was added for size comparison. The gel was dried and exposed to a PhosphoImager screen, and ImageQuant software was used to view and analyzed the gel data.

## 2-Aminopurine fluorescence assay

RNA containing 2-aminopurine in place of adenine at the indicated position (Dharmacon RNA Technologies)(Fig 6) was diluted to 500 nM in binding buffer (100 mM NaCl, 50 mM NaPO<sub>4</sub> pH 7.5, 100 µg/ml heparin, 3 mM βME). All reactions were performed at room temperature, in binding buffer, using a l-formate Jobin-Yvon Horiba Fluoromax fluorimeter and a 3 mm wide Spectrosil microcell cuvette (Starna Cells, Inc). Readings were taken without PPR5 added, and with indicated concentrations of PPR5 (5X was 2.5 µM PPR5, 10X was 5 µM PPR5) at 4 time points, immediately after addition of PPR5 (~30sec), 5 min, 10 min, and 15 min after addition of PPR5. The 2-aminopurine was excited at 315 nm, and spectra were collected from 320 to 420 nm. The fluorimeter slits were 2 nm, with an integration time of 0.1 seconds. Spectra collected with buffer, protein, and RNA without 2-aminopurine incorporated was used to subtract out background. The value at 370 nm was used to calculate relative fluorescence.

## Results

### The minimal PPR10 binding site spans 15 nucleotides

Previously we had localized a high affinity PPR10 binding site to a 29-nt segment of the *atpH* 5'-UTR (72). To better define the minimal region required to bind PPR10 with high affinity, we assayed its boundaries by performing binding assays with end-labeled RNA harboring the binding site that had been subjected to partial alkaline hydrolysis; the length of the shortest labeled RNAs capable of binding PPR10 defines the distance from the labeled end that is required for a high-affinity interaction. The results are shown in Figure 1A and summarized in Figure 1C. Analysis of 5' end-labeled RNA placed the 3' boundary required for high affinity PPR10 binding at position -29, with respect to the start of the *atpH* ORF. Analysis of 3' end-labeled RNA placed the 5' boundary at roughly -42, although RNAs with several additional nucleotides at the 5' end bind preferentially.

To validate and extend these conclusions, several synthetic RNAs were used in gel mobility shift assays (Figure 1B). An 18 nt RNA that lacks one nucleotide of the 3' boundary defined above failed to interact with PPR10, whereas all RNAs that include all the sequence within the 3' and 5' boundaries resulted in a high affinity interaction. Additional binding assays need to be done using synthetic RNA with the exact boundaries defined above to validate that these boundaries truly define the minimal ligand. 11/14 of the nucleotides in the minimal *atpH* binding site, as defined by the alkali hydrolysis binding assays described above, are shared in PPR10's second binding site, found in the *psaJ-rpl33* intergenic region. This striking conservation strongly suggests that most or all of the nucleotides within this RNA segment contribute to its specific interaction with PPR10.

The elution profile of recombinant PPR10 from a gel filtration column suggested that it might be a homodimer (72). To further address this possibility, we performed a stoichiometric binding assay in which the RNA was present at a concentration well above the  $K_d$ , and the fraction of RNA bound to protein was measured as a function of PPR10 concentration (Figure 2). The results show an inflection point at a PPR10:RNA ratio of ~2.5. This finding is consistent with the possibility that PPR10 binds RNA as a homodimer, although we cannot exclude the possibility that the high stoichiometry results from a population of inactive PPR10 molecules.

### **PPR10 protects its RNA ligand from 3' and 5' exonucleolytic cleavage *in vitro***

We showed previously that the PPR10 binding sites are found at the 5' or 3'-termini of those chloroplast RNAs that fail to accumulate in *ppr10* mutants (72). On that basis, we hypothesized that bound PPR10 blocks 3' and 5' exonucleases, thereby stabilizing adjacent RNA segments. To test whether bound PPR10 is sufficient to

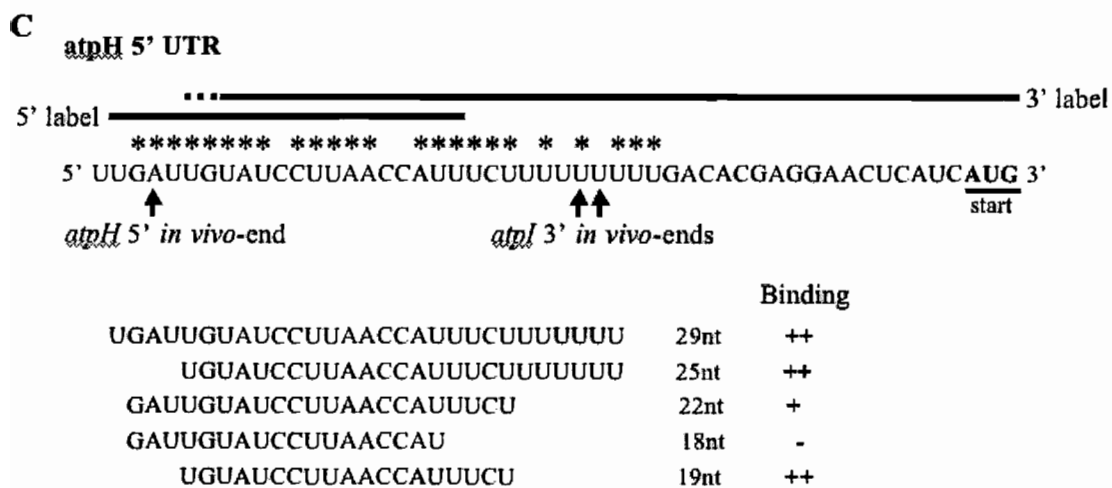
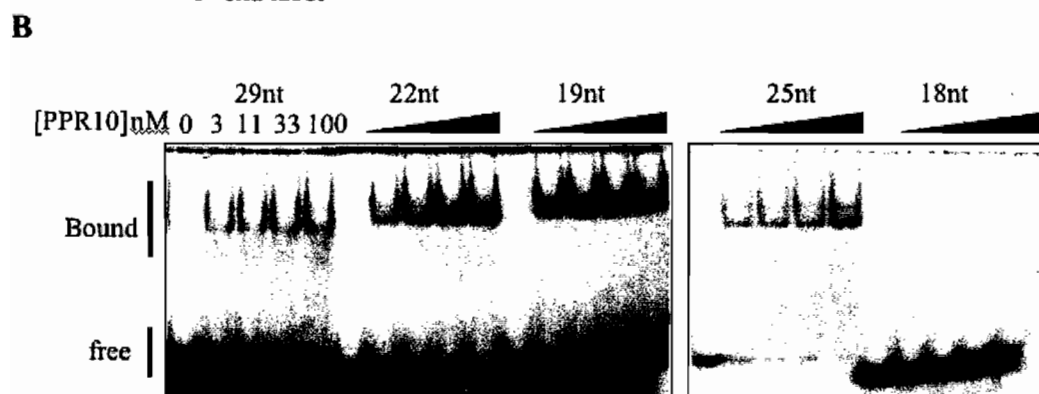
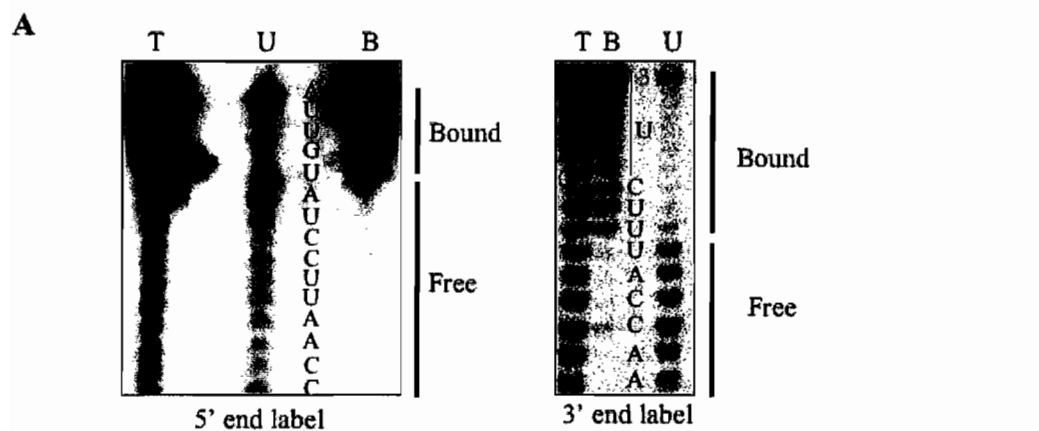


**Figure 1: The PPR10 RNA ligand.**

**(A)** Mapping the boundaries of sequences required for a high-affinity interaction with PPR10. The RNA shown in (C) was labeled at either its 5' or 3' end, subjected to partial alkali hydrolysis and used for gel mobility shift assays with PPR10. RNA was extracted separately from the gel regions containing unbound and bound RNA, and resolved on a denaturing polyacrylamide gel. The nucleotides assigned to each band were inferred based on their position from the labeled end. T- total hydrolyzed RNA. U- RNA that did not bind PPR10. B- RNA that bound PPR10.

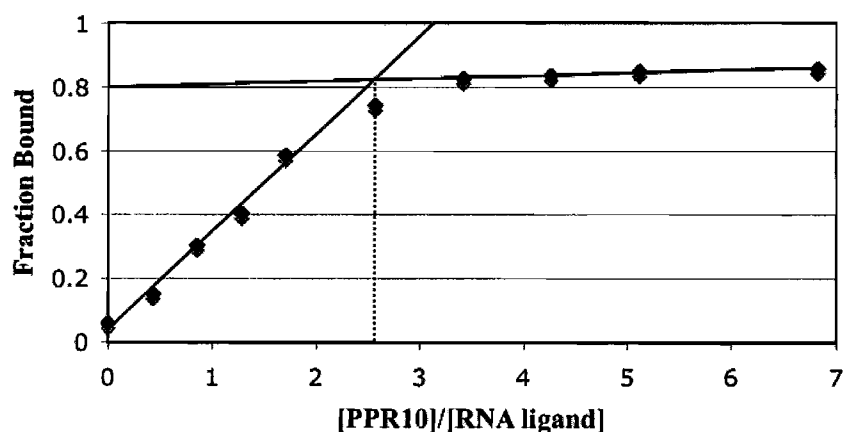
**(B)** Gel mobility shift assays, using the synthetic RNAs diagrammed in panel (C).

**(C)** Summary of data that define the minimal PPR10 binding site. The sequence of the synthetic RNA used for the boundary mapping experiment is shown at top; arrows annotate the major RNA termini defined by PPR10 *in vivo*, and asterisks annotate nucleotides conserved between the PPR10 binding sites in the *atpH* 5' UTR and the *psaJ* 3'UTR. The smallest end-labeled RNAs that bound well to PPR10 are indicated with bars; the 5' boundary is indicated with a dashed line, because of the gradient in apparent affinity observed as additional nucleotides in this region are included (see panel A, 3' end label). Smaller synthetic RNAs used for the gel mobility shift assays in panel B are shown below, annotated according to the degree to which they interact with PPR10.



block exonucleolytic RNA degradation *in vitro*, we performed *in vitro* assays with recombinant PPR10, synthetic end-labeled RNAs, and purified exonucleases (Figure 3). The *Synechocystis* polynucleotide phosphorylase (PNPase) was used as the 3'→5' exonuclease, as it is more easily expressed as a recombinant protein than is its chloroplast ortholog. Whereas the 5'-end labeled RNA alone was quickly degraded by PNPase, the addition of PPR10 inhibited degradation. (Figure 3B). The 3' termini that were stabilized

**Figure 2:** Stoichiometric binding assay with recombinant PPR10. Gel mobility shift assays were performed with a 5' end labeled synthetic 19 nt atpH 5' UTR RNA (Figure 1 C) at 100 nM concentration (Data not shown). The results were quantified by phosphorimaging. Linear trendlines were created in Excel using either the first 5 data points or the last 4 data points.



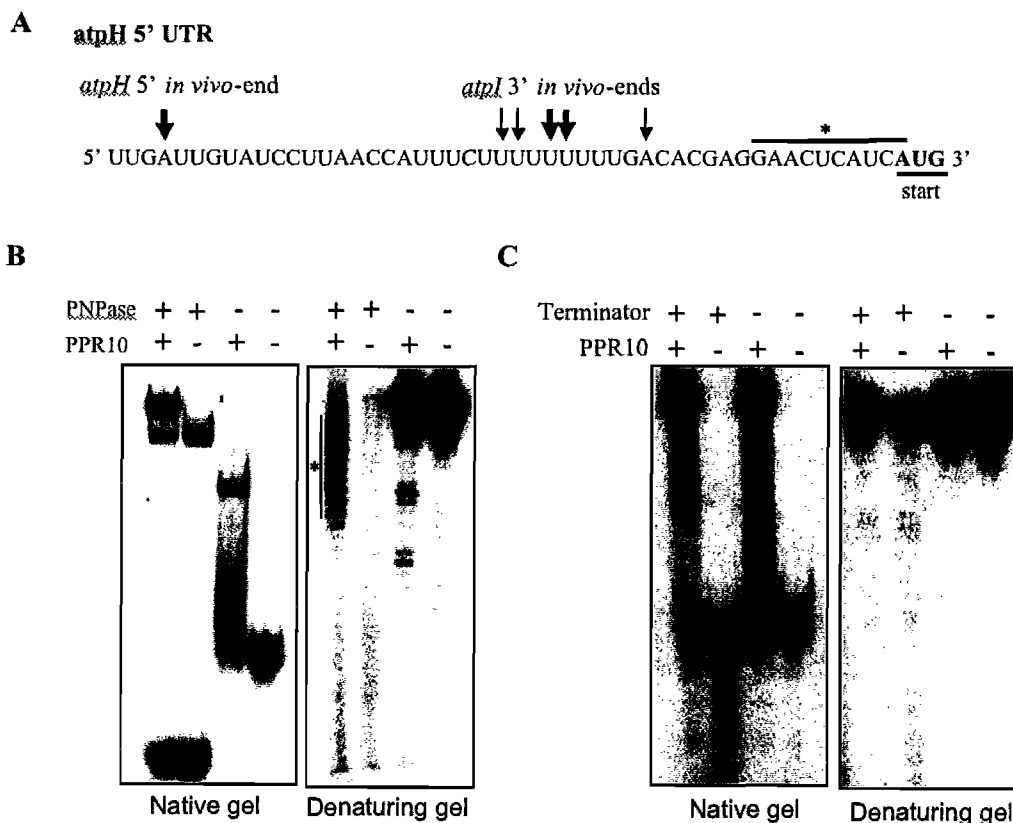
by PPR10 in this assay map ~10 nts downstream of the most abundant termini found *in vivo* (Figure 3A). There are several possibilities that can account for this. First, PNPase-mediated polyadenylation is believed to enhance processive RNA degradation through RNA secondary structures, but the reaction conditions used here were not optimized for polyadenylation activity. Second, PNPase is not the only 3'→5' exonuclease in

**Figure 3:** PPR10 protects against 3' and 5' exonuclease activity *in vitro*.

(A) Diagram of the RNA ligand used for nuclease-protection assays. The bar denotes the 3' ends that were protected from the PNPase assay shown in panel B.

(B) PPR10 protects RNA from *Synechocystis* PNPase *in vitro*. The left panel shows a gel mobility shift assay with the indicated proteins and the 5' end labeled RNA. The right panel shows a denaturing gel of the RNA recovered from the same reactions. The bar marks the termini of RNAs that were protected from PNPase digestion by PPR10. PPR10 and PNPase concentrations are 200 nM and 440 nM respectively.

(C) PPR10 protects RNA from a 5'→3' exonuclease. Terminator exonuclease (Epicentre Biotechnologies) and PPR10 were included in reactions with 3'-end labeled RNA, as indicated. The left panel shows a native gel mobility shift assay. The right panel shows a denaturing gel of the RNA recovered from the same reactions. PPR10 and terminator exonuclease concentrations are 200 nM and 1 μM respectively.



chloroplasts (reviewed in 7); thus, a different exonuclease could cooperate with the PNPase to generate the *in vivo* 3' end. Despite these caveats, these results suggest that bound PPR10 is sufficient to confer protection from 3'→5' exonuclease digestion. However, I plan to repeat this assay with chloroplast extract, using the conditions reported for efficient native PNPase activity (74).

The same RNA was labeled at its 3' end and incubated with a commercially available 5'→3' exonuclease. The addition of PPR10 fully protected the RNA from degradation (Figure 3C), indicating that bound PPR10 is sufficient to block access by 5'→3' exonucleases. Because the PPR10 binding site is near the 5' end of this RNA substrate, it is possible that bound PPR10 simply prevents the exonuclease from loading onto the RNA instead of blocking exonucleolytic progression. To address this possibility, I will repeat this assay with an RNA that has additional sequence upstream of the PPR10 binding site.

### **PPR10 binding releases the *atpH* ribosome binding site from a sequestering secondary structure**

Previously we had shown that the residual *atpH* mRNAs in *ppr10* mutants are translated less efficiently than their counterparts in normal plants, indicating that PPR10 binding simultaneously stabilizes *atpH* RNA and enhances its translation (72). In light of this observation, it is intriguing that the putative Shine-Dalgarno element for *atpH* translation is predicted to base pair with a portion of the PPR10 binding site (Figure 4A). Current data support the view that PPR tracts bind single-stranded but not double-stranded nucleic acids along their surface (71, 75, 76). Thus, we hypothesized that PPR10's interaction with the "anti-Shine-Dalgarno" element would prevent masking of the Shine-Dalgarno region, thereby facilitating ribosome recruitment.

To test this hypothesis we used ribonucleases T1 and V1 to probe the structure of the *atpH* 5' UTR in the presence and absence of PPR10 (Figure 4B). RNase T1 cleaves after guanosines, but only when they are in a single-stranded context; RNase V1 cleaves



regions in which the bases are stacked due to their presence in a double-stranded region or to other structural constraints. In the absence of PPR10, the three guanosine residues in the putative Shine-Dalgarno element were not cleaved by RNase T1, and the anti-Shine-Dalgarno element was efficiently cleaved by RNase V1. These results support the existence of the predicted RNA duplex in the majority of molecules. Addition of PPR10 caused a dramatic change in the digestion pattern. First, the guanosines within and a short distance upstream of, the Shine-Dalgarno element were now efficiently digested by RNase T1, indicating a substantial increase in their single-stranded character. Second, RNase V1 ceased to cleave the anti-Shine-Dalgarno region; this could be due either to direct protection by PPR10 or to a PPR10-induced loss of the RNA duplex. Finally, PPR10 binding increased RNase V1 sensitivity at several positions 3' to the PPR10 binding site. PPR10 apparently induces the stacking of these bases, but details of these changes cannot be inferred from these data. It is intriguing, however, that a similar enhancement of RNase V1 cleavage was observed adjacent to RNA bound by PPR5 (see below).

These data show that PPR10 binding induces a rearrangement of the RNA in the *atpH* 5' UTR. The PPR10-induced rearrangement would be anticipated to enhance translation regardless of whether the putative Shine-Dalgarno site indeed has ribosome binding activity, as initiating ribosomes interact with ~30 nucleotides of single-stranded RNA centered on the start codon (77). Taken together, these results support a model in which PPR10 captures its binding site in the *atpH* 5'UTR in single-stranded form, thereby increasing the single-stranded character of the *atpH* ribosome binding region and facilitating ribosome binding (Figure 4C).

### **The PPR5 binding site is complex and includes discontinuous RNA segments**

To understand general features of PPR/RNA interactions, it is necessary to analyze multiple examples. Thus, a second PPR protein, PPR5, was analyzed in parallel with PPR10. Previously we had determined that the PPR5 binding site resides within a 50 nt segment of the group II intron in pre-trnG-UCC (71). When PPR5 binds to this site

*in vivo*, it stabilizes the unspliced precursor by blocking an endonucleolytic cleavage (48). In addition, PPR5 appears to enhance splicing itself, as the ratio of spliced-to-unspliced *trnG* RNA is substantially reduced in hypomorphic *ppr5* mutants. A direct role for PPR5 in splicing is consistent with the fact that its binding site contains several sequence elements that are important for group II intron splicing: Exon Binding Site 1 (EBS1),  $\alpha'$ , and  $\delta$  (see Figure 5A). In order for splicing to occur, each of these sites must pair with complementary sequences found elsewhere (IBS1,  $\alpha$ , and  $\delta'$ , respectively) (reviewed in 78). The EBS1 and  $\delta$  elements in this intron are unusual, in that they are predicted to be sequestered in a stable RNA hairpin (Figure 5A); formation of this structure is supported by the ribonuclease-sensitivity data described below. Thus, we hypothesized that PPR5 binding may enhance splicing by influencing the structure of this RNA (71).

To understand how PPR5 could influence the splicing of its group II intron ligand, we initially defined its binding site more precisely by using assays analogous to those described above for PPR10. The PPR5 analysis was more complex than that of PPR10 for two reasons. First, previous data suggested that PPR5 interacts with discontinuous RNA segments ((71), and this possibility was supported by the additional results described below. Second, the RNA sequence harboring the PPR5 binding site has the capacity to form several alternative structures, with the favored structure changing as various segments are removed (data not shown). Because PPR5 binds preferentially and possibly solely to single stranded RNA (71), failure of a deletion construct to bind to PPR5 could potentially be due to sequestration of PPR5 recognition elements within an RNA duplex.

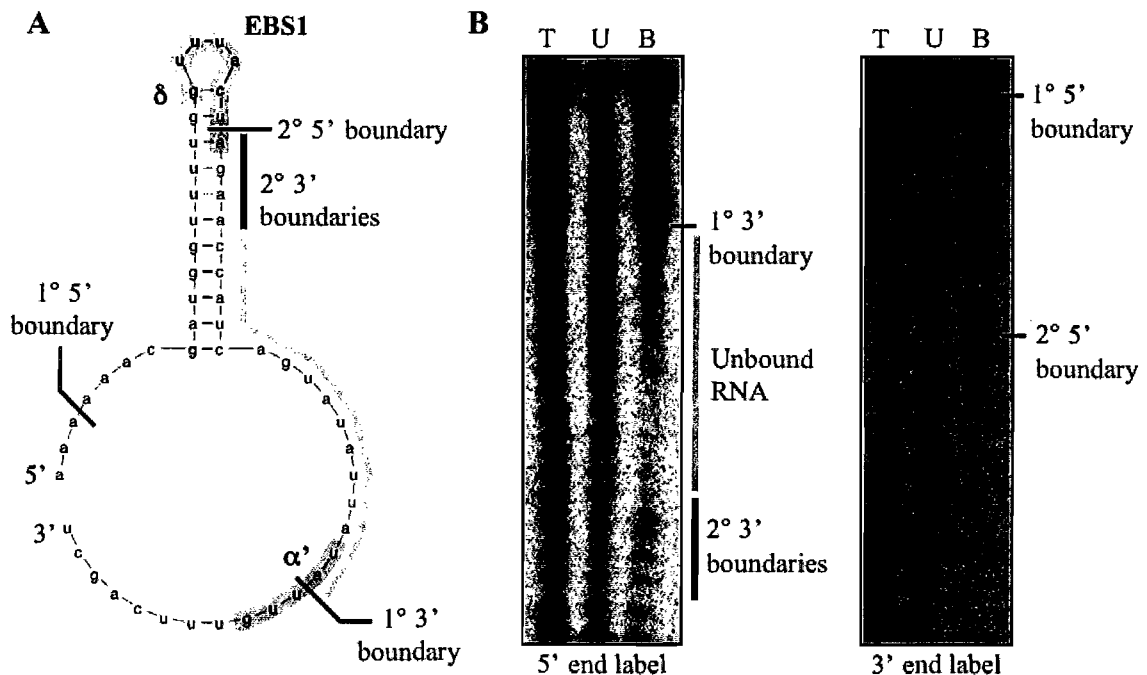
When a partial alkali hydrolysis binding assay with 5' end-labeled RNA was performed, the shortest RNA that bound with high affinity to PPR5 terminated two nucleotides into the  $\alpha'$  element, suggesting that recognition determinants for PPR5 lie within or just upstream of  $\alpha'$  (see 1° 3' boundary in Figures 5A and B). Although binding was lost for molecules ending in the single-stranded region upstream of  $\alpha'$  (gray bar in Figure 5B), weak binding was detected after further truncation to remove the base



**Figure 5:** Mapping the boundaries of sequences required for a high-affinity interaction with PPR5.

(A) Predicted secondary structure of the region harboring the PPR5 binding site. Elements involved in group II intron splicing (EBS1,  $\delta$ , and  $\alpha'$ ) are marked. The boundaries mapped in the experiments shown in (B) are indicated.

(B) Partial alkali hydrolysis binding assay, using 5' or 3'-end labeled RNA. The nucleotides assigned to each band were inferred based on their position from the labeled end and by comparison to a nuclease T1 ladder. T- total hydrolyzed RNA. U- RNA that did not bind PPR5. B- RNA that bound PPR5.



of the 3' side of the hairpin as these molecules were depleted from the unbound fraction, and enriched in the bound fraction (see 2° 3' boundary in Figure 5A and B). Therefore, PPR5 can interact with molecules that end within the distal side of the stem, albeit with lower affinity than with the full-length 50-mer. That deletion of the distal side of the stem was required to reveal this secondary interaction suggested that sequences on the 5' side of the stem are important for PPR5 binding, and that these are masked when the stem is intact.

A partial alkali hydrolysis binding assay with 3'-end labeled RNA revealed that truncation of the 5' end past position 3 reduced binding dramatically (see 1° 5' boundary in Figure 5 A and B). Thus, these boundary mapping experiments implicated sequences both 5' and 3' to the stem as being important for PPR5 recognition, consistent with gel mobility shift data reported previously (71).

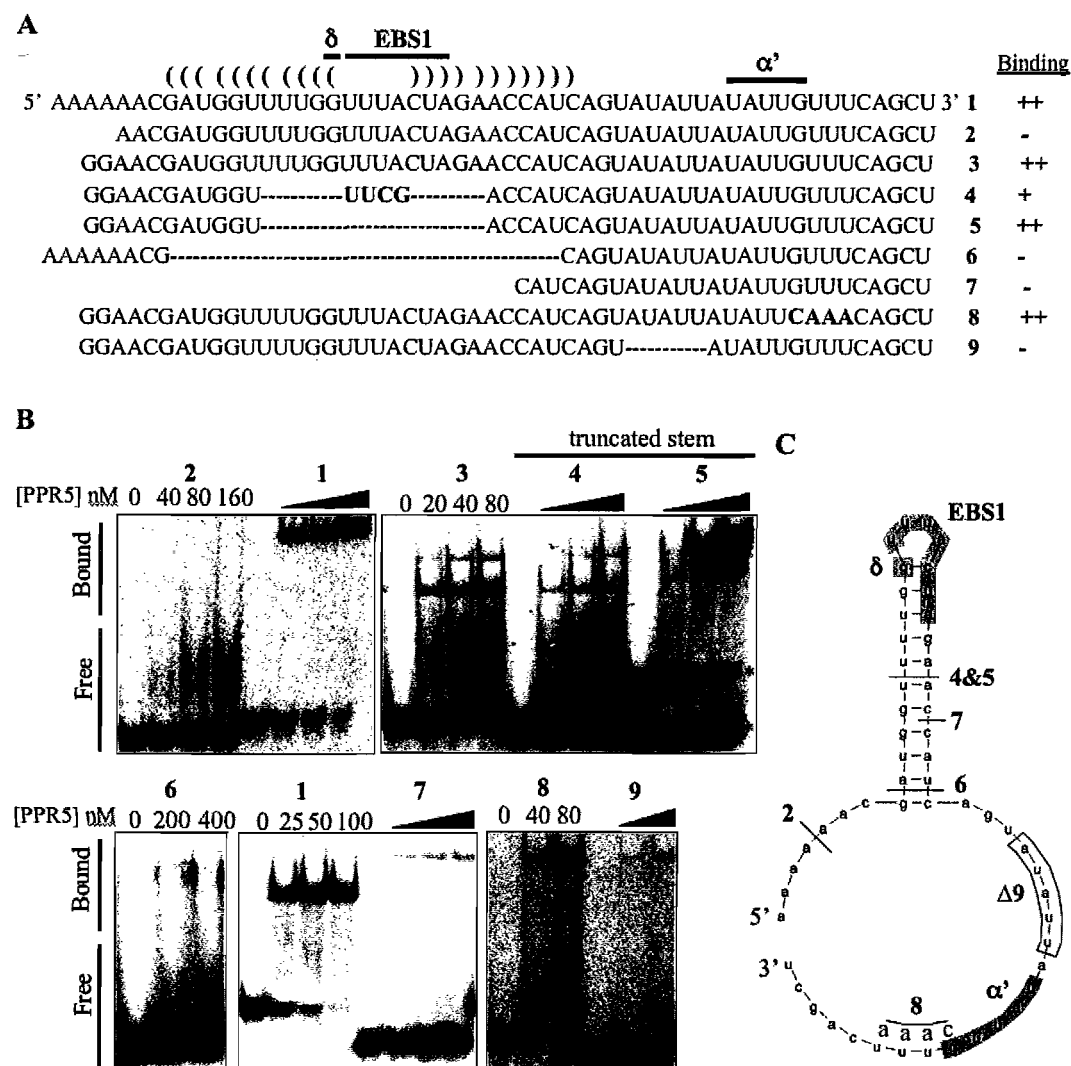
Gel mobility shift assays with synthetic oligonucleotides (Figure 6) confirmed that RNA sequences on both sides of the hairpin are required for a high-affinity interaction with PPR5. For example, removing the four adenine residues at the 5' end caused a dramatic decrease in binding (Figure 6B, construct 2), as did deletion of five nucleotides within the 3' single-stranded region (Figure 6B, construct 9). To determine whether sequences within the stem contribute to binding affinity, various stem truncations were assayed. Previously we showed that deletion of the EBS1 element did not disrupt binding (71). An RNA lacking the distal half of the stem (construct 5) maintained considerable binding activity. This RNA apparently adopts two structures that migrate differently through a native gel (see asterisks); only the more slowly migrating conformer bound PPR5, as only this form was depleted as PPR5 concentrations increased. That a high affinity interaction with PPR5 requires some invasion of the stem was suggested by the fact that stabilizing the truncated stem with a terminal tetraloop decreased its interaction (compare constructs 4 and 5). Furthermore, removal of the entire stem eliminated binding (Figure 6B, construct 6), strongly suggesting that recognition determinants reside within the stem itself. Indeed, the

**Figure 6:** The PPR5 RNA ligand.

**(A)** Alignment of the PPR5 ligand region and truncations used for gel mobility shift assays. Exon Binding Site 1 (EBS1), delta ( $\delta$ ) and alpha prime ( $\alpha'$ ) are labeled. Predicted stem denoted by parenthesis. Sequence 4 contains a tetraloop (UUCG) that promotes stem formation. PPR5 binding affinity indicated by ++ (high affinity), + (moderate), and - (no binding).

**(B)** Gel mobility shift assays showing PPR5 binding to the truncations of the trnG RNA shown in A. Sequence 5 apparently adopts two conformations that migrate differently in the gel (\*).

**(C)** Diagram of the region of the trnG intron to which PPR5 binds. Lines indicate which part of the RNA molecule was removed or altered (in the case of 8) resulting in the constructs in A). Sequences that were bound by PPR5 are labeled in gray whereas sequences that did not bind are labeled in black. EBS1,  $\delta$ , and  $\alpha'$  are in gray boxes.



boundary-mapping data using 5' end labeled RNA revealed a 2° interaction site upon removal of the 3' end of the stem (see above), implicating nucleotides near the 5' end of the stem as contributing to PPR5 binding.

Previously, we had reported that elimination of sequences downstream of  $\alpha'$  prevents the binding of PPR5 (71). In this study we found that a GUUU to CAAA substitution just downstream of  $\alpha'$  greatly increases PPR5 binding (Figure 6B, construct 8). These data suggest that PPR5 may be interacting with sequences 3' of the  $\alpha'$  site as well as sequences 5' of this site.

Taken together, these results support the view that PPR5 recognizes nucleotides that are discontinuous in the primary sequence; these include the 5' single-stranded region, one or several nucleotides on the 5' side of the stem base, the single-stranded region on the 3' side of the stem adjacent to  $\alpha'$ , and perhaps nucleotides on the 3' side of  $\alpha'$ . That deletion of the entire stem eliminates binding is an important observation, as this provides evidence that PPR5 invades the stem, providing a plausible mechanism by which it could influence the stability of the hairpin, and thus the efficiency of splicing. It will therefore be important to firmly establish the location of PPR5 recognition determinants at the base of the RNA stem. To test the notion that the nucleotides on the 5' side of the stem contribute to a high affinity interaction with PPR5, I plan to test several additional constructs. For example, I will test the binding activity of an RNA harboring nucleotides 1-11, fused directly to nucleotides 33 through 50.

### **PPR5-induced changes in RNA structure suggest mechanisms by which PPR5 enhances splicing**

Group II intron splicing requires the EBS1,  $\alpha'$ , and  $\delta$  elements within the intron to base pair with their complements found elsewhere. Consequently, these elements are found in a single stranded context in the vast majority of group II introns (reviewed in 78). In this context, the apparent sequestration of EBS1 and  $\delta$  in the PPR5 binding region within the trnG-UCC intron (Figure 5) are striking. The binding data suggested that PPR5 might destabilize this hairpin (and thereby activate splicing) by invading the 5' side of the

stem. To address how PPR5 influences the structure of its RNA ligand, we used ribonucleases T1, V1 and R1 to probe RNA structure in the presence and absence of PPR5 (Figure 7). RNase T1 cleaves after single-stranded guanosines, RNase R1 cleaves single-stranded pyrimidines, and RNase V1 cleaves stacked or double-stranded regions. The results obtained in the absence of PPR5 (Figure 7A) provided support for the predicted stem-loop structure. For example, RNases R1 and T1 cleaved the region between the predicted stem and  $\alpha'$ , but did not cleave within the predicted stem (lanes 5-8 and 13-16). RNase V1, in contrast, cleaved many of the positions predicted to reside within the stem (lanes 9-12). The EBS1 was cleaved weakly by RNase R1 (lanes 13 and 14), but not at all by RNase V1 (lanes 9-12).

When PPR5 was bound to the RNA prior to the ribonuclease treatments, the cleavage patterns changed in several interesting ways. For example, the single-stranded region upstream of  $\alpha'$  became less sensitive to cleavage by RNase R1 (lane 14), suggesting an interaction between PPR5 and these nucleotides (see dark gray bar in Figure 7). Indeed, deletion of nucleotides in this region caused a dramatic decrease in PPR5 binding (see construct 9 in Figure 6). The G residue at the 5' base of the stem became susceptible to cleavage by RNases T1 and R1 (see G\* lanes 5-8 and 13-16 in Figure 7), suggesting that PPR5 binding releases this nucleotide from an RNA duplex. The most dramatic effect, however, concerned the  $\alpha'$  region, which became hypersensitive to all three nucleases upon PPR5 binding (lanes 5-16). Enhanced cleavage by RNases T1 and R1 suggested an increase in single-stranded character, yet the increased sensitivity to RNase V1 indicated increased base-stacking or base-pairing. These observations suggested that PPR5 binding constrains the structure of the RNA in the  $\alpha'$  region, such that the bases are single-stranded but stacked. Curiously, however, the  $\alpha'$  residues whose RNase T1 sensitivity increased upon PPR5 binding are not adjacent to G residues, and were not susceptible to RNase T1 cleavage even in the fully denatured RNA (see T1 marker, lane 1, in Figure 7). A profound change in the structure of these nucleotides is further supported by the fact that the G residue within  $\alpha'$  site becomes sensitive to RNase H cleavage in the presence of PPR5 (lanes 17 and 18), yet

**Figure 7:** Ribonuclease sensitivity assay of RNA structure in the absence and presence of PPR5.

**(A)** The RNA shown in panel B was labeled at its 5' end, incubated in the absence (-) or presence (+) of PPR5, and then treated with RNase T1, V1, or R1. Two concentrations of each nuclease were tested, with the left pair of lanes in each instance representing the higher concentration. The T1 marker was generated by treating the same RNA with RNase T1 after heating and snap-cooling to minimize secondary structures. The OH marker is a partial alkali hydrolysis, to mark the positions of consecutive nucleotides. PPR5 incubated under the same conditions used for the nuclease treatment did not cause any RNA cleavage (lane 4). G\* refers to a residue that becomes susceptible to RNase T1 and RNase R1 after PPR5 binding. Other features referred to in the text are coded with bars to the right, and summarized in panel B.

**(B)** Summary of structure probing data. EBS1 and  $\alpha'$  sites are outlined and indicated by black bars in (A). G\* residue that becomes susceptible to RNases T1 and R1 with PPR5 addition is encircled. Stem structure is shaded gray and indicated by gray bars in (A). Region that is protected from RNase R1 cleavage when PPR5 is added is outlined and shaded gray, and indicated by a dark gray bar in (A).



RNase H is believed to be specific for RNA found in the context of an RNA/DNA hybrid. Taken together, these results show that PPR5 binding causes the  $\alpha'$  site to become hypersensitive to nucleases, including nucleases that would not ordinarily recognize those particular sequences. These observations suggest that PPR5 changes the architecture of this region of the RNA in an unusual way, and that this might be important for presenting the  $\alpha'$  element during the splicing reaction.

The nuclease-sensitivity data offered hints that PPR5 binding may reduce the double-stranded character of EBS1 and its flanking sequences. As noted above, the G residue at the 5' end of the stem becomes susceptible to RNase T1 cleavage upon PPR5 binding. In addition, a subtle but reproducible enhancement of RNase R1 cleavage of EBS1 sequences is induced by PPR5 (Figure 7, bar labeled EBS1). To gain further insight into this possibility, the formation of the stem was probed by substituting a fluorescent analog of adenine, 2-aminopurine (2-AP), for one of the adenines on the distal side of the stem (Figure 8A). 2-AP fluorescence is high when it is unstacked but greatly decreases when it is stacked due to its residing in an RNA duplex or other structural constraints (79). Addition of PPR5 to the 2-AP modified RNA resulted in an increase of fluorescence, indicative of reduced base stacking in the stem (Figure 8B). PPR10, which does not bind with high affinity to the PPR5 binding site, also induced some increase in fluorescence, although to a much lesser extent. This is consistent with the fact that PPR10 binds weakly to this RNA under the reaction conditions used (data not shown). Initially, the enhanced fluorescence caused by PPR5 in comparison with PPR10 increased with time. This could be due to different binding kinetics, or could be a result of nucleases in the PPR5 preparation. I plan on resolving these possibilities at a later date by isolating the RNA directly after determining the fluorescence and assaying for degradation.

These uncertainties notwithstanding, the body of nuclease-sensitivity and 2-AP fluorescence data support the idea that PPR5 binding results in an increase in the single-stranded character of the RNA duplex harboring EBS1 and  $\delta$ . As PPR5 binds to single stranded but not double stranded RNA (71) and clearly requires nucleotides within the

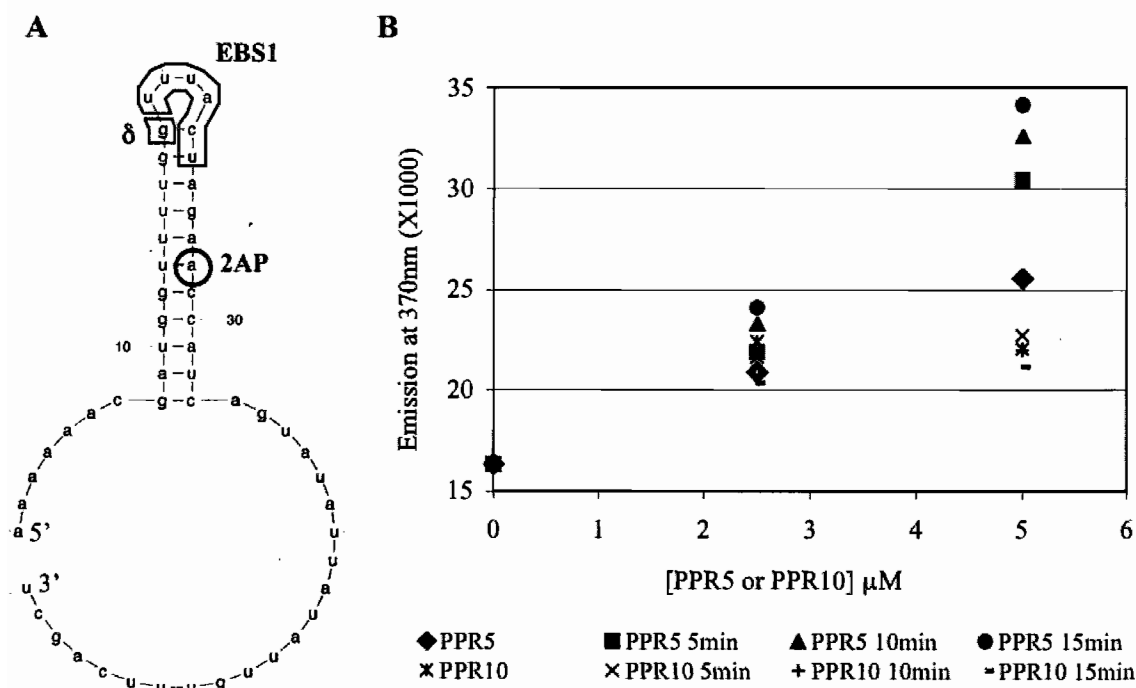


stem for a high-affinity interaction (Figure 6, construct 6), a reasonable hypothesis is that PPR5 binding nucleates in the single-stranded regions flanking the stem, but ultimately invades the stem base. This may enhance the splicing of the *trnG* intron, by increasing the accessibility of the EBS1 and  $\delta$  elements to their RNA partners.

**Figure 8:** PPR5 causes an increase in 2-aminopurine fluorescence in its ligand, indicating unfolding of the RNA stem.

(A) The *trnG* RNA ligand of PPR5 with Exon Binding Site One (EBS1),  $\delta$ , and the position of 2-aminopurine labeled.

(B) fluorescence emission at 370nm for RNA [.5 $\mu$ M] mixed with PPR5 or PPR10. Readings were taken either right away, or after 5, 10, and 15 minutes.



## Discussion

In this study we have defined the PPR10 and PPR5 binding sites to high resolution, and we showed that both proteins profoundly influence the structures adopted by the RNA flanking their binding sites. These results have broad implications regarding mechanisms by which PPR proteins recognize RNA and influence downstream functions. PPR proteins have been implicated in a variety of processes, including RNA splicing, RNA editing, RNA cleavage, RNA stabilization, and translation control. Because these functions appear to be diverse, it has often been suggested that PPR proteins serve as adaptors to recruit various effector proteins to specific RNA sites. However, with the notable exception of PPR proteins involved in RNA editing, experimental evidence to support this view is lacking. Our results suggest an alternative possibility: that most functions attributed to PPR proteins- particularly those consisting largely of “pure” PPR repeats - result as a passive consequence of their sequence-specific binding to long tracts of single-stranded RNA. Below we discuss evidence that the PPR/RNA interaction interface is unusually long in comparison with those mediated by most RNA binding motifs, and that this activity in itself could account for many of the dramatic and diverse effects of PPR proteins on organellar RNA metabolism.

### **Features of the PPR10 binding site suggest that PPR10 binds RNA along an unusually long RNA/protein interface**

Several observations support the idea that PPR10's RNA interaction surface is substantially longer than that of typical RNA binding proteins. Most RNA binding proteins contain several globular RNA binding domains, such as the RRM or KH domain, each of which contacts ~2-5 nucleotides. The combinatorial action of several domains and their variable orientation with respect to one another can mediate the recognition of specific RNAs based on a combination of sequence and structure (80-82). In contrast, the minimal RNA segment required for a high affinity interaction with PPR10 spans ~15 nt, with its *in vivo* footprint (i.e. the RNA protected by PPR10 from

exonucleases *in vivo*), substantially larger, at ~25 nt. The extremely high conservation of the nucleotides within this RNA segment provides evidence that most of its nucleotides contribute to binding affinity. Thus, PPR10's second binding site, which maps in the *psaJ-rpl33* intergenic region, has only three nucleotide differences, and even these small differences are associated with a substantial decrease in binding affinity (72). Furthermore, the sequence of the 25 nts within PPR10s *in vivo* footprint in the *atpH* 5' UTR is almost identical in monocot and dicot plants (e.g. maize and spinach differ at only one of 25 positions). Although no other PPR proteins have been analyzed in this level of detail, several compelling observations support the view that the PPR protein HCF152 likewise has an extensive *in vivo* footprint and that its binding site is extremely highly conserved between monocots and dicots (72). These data, albeit still limited, suggest that an extensive RNA/protein interface along which most contiguous nucleotides interact with the protein is the norm for PPR proteins harboring long tracts of canonical PPR repeats.

That long PPR tracts have an extensive RNA interaction surface is consistent with structural predictions. The PPR motif is closely related to the TPR motif, a 34 amino acid repeating unit that generally mediates protein-protein interactions (10, 83). TPR tracts adopt a helical repeat solenoid structure (83-85), with each repeat forming a pair of helices, and consecutive repeats stacking to form a broad substrate-binding surface. It is anticipated that PPR tracts likewise form helical repeat solenoids, although structural data remain very limited (71). This is an atypical structure for a nucleic acid binding protein, but there is precedent in the PUM-Homology Domain (PUM-HD). The PUM-HD defines the "PUF" protein family, whose members regulate gene expression in eukaryotes by binding specific 3' UTRs and influencing RNA stability or translation (reviewed in 86). Structural analyses revealed an unusual mechanism for RNA recognition: the PUM-HD consists of eight helical repeating units; consecutive repeats stack to form an RNA binding surface, with each repeat recognizing a single RNA base (87). Our results support the view that PPR tracts likewise bind single-stranded RNA parallel to the axis of stacked alpha helices. Whereas the PUM-HD always consists of eight repeats and binds

an ~8 nt core element, the number of repeats in PPR proteins is highly variable and generally greater, with 20 repeats commonly observed. Thus, according to this model of PPR/RNA recognition, the length of the PPR/RNA interaction surface is limited only by the number of PPR motifs.

PPR10 contains 16 canonical PPR motifs that are preceded by two additional repeats that have more TPR character (72). That PPR10's minimal RNA ligand spans 14-16 nt is intriguing in light of its 16 PPR motifs, as it suggests that each nucleotide may be recognized by a single PPR motif. However, prior observations suggested that recombinant PPR10 forms homodimers (72), and the stoichiometric-binding assay presented here supports this view, in that two molecules of PPR10 appear to bind to each *atpH* 5'UTR. One possible explanation for this apparent discrepancy is that one monomer binds in a sequence-specific manner to the minimal binding site whereas the other binds to adjacent regions in a sequence-non-specific fashion. This view is consistent with the finding that PPR10's *in vivo* footprint is significantly longer than its minimal binding site.

The PPR5/RNA interaction is considerably more complex, and therefore is less informative regarding the relationship between the number of PPR motifs and the number of nucleotides recognized. PPR5 recognizes two non-contiguous RNA segments within a 50-nt RNA sequence, and this RNA has a propensity to fold into various stable RNA structures. Our results indicate that PPR5 can bind to either the 5' or 3' portion of the 50-mer, but that it binds with highest affinity when both regions are present in the same molecule. In aggregate, our results lead us to favor a model in which two molecules of PPR5 bind to each 50-mer, one interacting with the 5' single-stranded region and invading the RNA duplex, the other interacting in the single-stranded region between the duplex and alpha'. We further speculate that two PPR5 monomers bind to this RNA cooperatively, as recombinant PPR5 did not dimerize, and gel mobility shift assays did not provide evidence for complexes of varying mobility as the PPR5 concentration was increased (71). Additional experiments will be required to fully understand these interactions.

**Site-specific barrier and RNA remodeling functions of PPR5 and PPR10: implications for the mechanisms by which PPR proteins mediate downstream effects**

Genetic data have implicated proteins composed virtually entirely of PPR motifs in diverse functions, including RNA cleavage, RNA stabilization, translational control and group II intron splicing. Thus, it has often been thought that they serve as sequence-specific adapters whose sole function is to recruit effector proteins to appropriate RNA sites. Our findings with PPR5 and PPR10 suggest an alternative view: that the unusual features of the PPR/RNA interface can directly result in most or all of the *in vivo* functions attributed to “pure” PPR proteins (i.e. those composed almost entirely of canonical PPR motifs) without the involvement of accessory factors. We propose that: (i) long PPR tracts sequester an extended segment of single-stranded RNA; (ii) that this activity makes them particularly effective at blocking access to their RNA ligands by other proteins and at remodeling adjacent RNA structures; and (iii) that these two effects are sufficient to account for the many biological functions attributed to proteins of this nature.

Our results with PPR5 and PPR10 illustrate how a pure PPR protein can, on its own, enhance the splicing, translation, or stability of specific RNAs, and can appear to enhance site-specific RNA processing events. We showed previously that PPR10 is required for the accumulation of those RNAs harboring its binding site at either their 5' or 3' end, suggesting that PPR10 serves as a barrier to exonucleases intruding from either direction (72). Here we present evidence that PPR10 is sufficient to block exoribonucleases *in vitro*. Additional genetic data support the idea that a blockade to 5' → 3' degradation is a common function of PPR proteins (e.g.88). Recently, a moss PPR protein was shown to stabilize its RNA ligand against 3' → 5' exonucleases *in vitro* (89). Finally, PPR5 stabilizes the trnG-UCC precursor *in vivo* against an inactivating endonucleolytic cleavage (48, 71). Together, these results strongly suggest that the ability to block ribonuclease access to its RNA ligand is an intrinsic activity of long PPR tracts. Genetic data have provided evidence that some PPR proteins repress the translation of specific organellar mRNAs. This activity can likewise be accounted for by a passive

“barrier” activity, as an extensive interaction with an RNA segment that includes nucleotides required for interaction with initiating ribosomes would surely inhibit translation initiation.

In addition to blocking access of bound RNA to other proteins, it is anticipated that an interaction with a long PPR tract will likewise block interaction of an RNA segment with complementary RNA sequences. This, in turn, can influence RNA folding in a manner that can account for the ability of pure PPR proteins to activate translation, splicing, and even RNA cleavage. Results presented here for PPR5 and PPR10 provide evidence for this type of RNA remodeling activity. PPR10 binding enhances the translation of the adjacent *atpH* open reading frame *in vivo*. The PPR10 binding site includes sequences that are complementary to the putative Shine-Dalgarno element for *atpH* translation. We show here that PPR10 binding releases the Shine-Dalgarno element from sequestration in an RNA duplex, providing a plausible mechanism to explain its translation enhancing effects. An analogous mechanism can account for genetic data suggesting a translation activating function for other PPR proteins, with no need to invoke active recruitment of components of the translation machinery.

PPR5 is one of several PPR proteins that have been shown to enhance the splicing of group II introns *in vivo*. We believe the mechanism by which PPR5 promotes *trnG* intron splicing mirrors the mechanism by which PPR10 promotes translation. PPR5 binds RNA that is adjacent to the critical splicing elements EBS1,  $\delta$ , and  $\alpha'$ , which need to base pair with their complementary sequences for splicing to occur. Without PPR5, EBS1 and  $\delta$  are sequestered in a stem loop structure, and the presence of PPR5 destabilizes this structure. We propose that the PPR5 binding site includes several nucleotides at the base of the stem loop, and that PPR5 binding promotes the unfolding of the stem by capturing its RNA ligand in a single stranded conformation. In addition, PPR5 induces an unusual spectrum of nuclease hypersensitivity within the  $\alpha'$  sequence: PPR5 binding enhances cleavage by both single-strand and double-strand “specific” ribonucleases (RNAses T1 and V1), by RNase H in the absence of an RNA/DNA duplex, and by RNase T1 at residues other than guanines, its normal substrate. These results suggest that PPR5

distorts the RNA in close proximity to its binding site, possibly in a manner that makes  $\alpha'$  more accessible for pairing with its  $\alpha$  complement.

In summary, the ability of PPR5 and PPR10 to influence the structure adopted by adjacent RNA segments can explain their ability to activity translation and splicing, respectively. An analogous mechanism may account not only for other instances in which pure PPR proteins enhance the translation or splicing of specific RNAs, but also for the ability of some PPR proteins to enhance endonucleolytic processing at specific sites. For example, the binding of a PPR protein to an intergenic region on a polycistronic RNA could influence the adjacent RNA structure, and thereby expose a segment of RNA with features that make it susceptible to cleavage by generic endonucleases. We proposed previously that RNases E and J are primarily responsible for the endonucleolytic cleavage events that initiate both RNA processing and RNA decay in chloroplasts. The bacterial orthologs of these enzymes cleave AU-rich RA segments found in an unstructured context. Thus, a PPR binding site that includes an AU rich RNA segment can be anticipated to stabilize nearby RNA, whereas a PPR binding site adjacent to an AU rich RNA segment could enhance its accessibility to these nucleases by minimizing local RNA structure.

Our model that many functions attributed to PPR proteins are a passive consequence of the unusually extensive protein/RNA interface that is predicted for these proteins is limited to those PPR proteins that lack additional domains. In fact, many PPR proteins in plants include one of the accessory domains denoted as E, E+, or DYW. These proteins are involved in RNA editing, an activity that certainly requires a catalytic activity. Furthermore, the PPR tracts in such proteins are variants of the regular repeating array of tandem PPR motifs found in proteins such as PPR5 and PPR10; this variant organization, designated "PLS", is likely to interact with RNA in a less regular way, possibly of a "looser" nature. Although a recruitment function need not be invoked to explain most of the genetic data obtained for pure PPR proteins, our model does not preclude the possibility that some pure PPR proteins do interact with other proteins; indeed the homodimerization of PPR10 provides evidence for a protein-protein

interaction surface on this protein.

Genetic analysis has identified many PPR proteins that have diverse functions related to RNA binding. We believe that many of these functions may be mediated by the PPR proteins ability to promote single-stranded RNA conformation through, and adjacent to, its binding site. Biochemical approaches that identify the specific binding sites of these proteins, and analysis of these sites, will be essential in determining how prevalent this model for PPR protein function is.



## CHAPTER IV

### CONCLUSIONS AND FUTURE DIRECTIONS

#### Conclusions

This dissertation investigates the nuclear control of gene expression in the chloroplast. The plant nucleus encodes many families of proteins that are targeted to the chloroplast where they regulate expression of the chloroplast genome. Some of these proteins are descendent from cyanobacterial proteins that may have had similar function prior to endosymbiosis. Many of these proteins are host innovations that evolved to accommodate the needs of a changing chloroplast genome. The proteins discussed here, WHY1, PPR5, and PPR10, are examples of the latter group of host-derived proteins.

Chapter II discusses WHY1, a chloroplast targeted, nuclear encoded protein that binds to both single-stranded DNA and single-stranded RNA. WHY1 binds with specificity to the *atpF* group II intron and promotes *atpF* splicing. However, *why1* mutant plants show a strong albino phenotype that cannot be accounted for by this splicing defect alone. Examination of *why1* mutant transcripts revealed a severe loss of 23S and 4.5S ribosomal RNAs suggesting that WHY1 is involved in the biogenesis of the large ribosomal subunit. WHY1 does not appear to bind directly to ribosomal RNAs, or mature ribosomes, leading us to conclude that WHY1's influence on ribosomal biogenesis is most likely indirect.

WHY1 also binds DNA throughout the chloroplast genome and has a strong preference for single-stranded DNA. Though other groups have proposed functions for WHY1 binding to DNA in the nucleus (26, 27, 29), the significance of this property in the chloroplast still remains to be elucidated. WHY1 does not appear to be involved in

nucleoid DNA replication or global transcription as chloroplast transcript and DNA abundance in *why1* mutants is similar to that of relevant controls. WHY1 coimmunoprecipitates all chloroplast DNA sequences suggesting the WHY1/DNA interaction is not sequence-specific, although the possibility that WHY1 binds with sequence-specificity to a sequence represented throughout the entire chloroplast genome cannot be excluded.

Chapter III discusses the PPR family of proteins. PPR proteins are involved in many diverse RNA-related functions. However, most PPR proteins lack obvious catalytic domains. Because of this, they are often proposed to be sequence-specific adaptors that recruit effector proteins to appropriate RNA sites. We have shown here, that the unique PPR/RNA interaction surface can directly explain many of the functions attributed to PPR proteins. The results presented in this dissertation provide three examples: RNA stabilization, translational activation, and RNA splicing.

The previously accepted model of RNA processing in the chloroplast suggested that mature RNA termini are defined by site-specific endonucleolytic cleavage of polysistronic transcripts (7). According to this model, sequence-specific RNA binding proteins, like PPR proteins, define mature RNA ends by recruiting endonucleases to specific cleavage sites. Data presented here, as well as in our previous investigations, suggest an alternative model in which PPR proteins define mature RNA termini by acting as a site-specific barrier to exonucleolytic cleavage (72). In this model endonucleases cleave polysistronic transcripts at exposed (ribosome free) AU rich regions. Exonucleases subsequently degrade the RNA from the cleavage site until their progression is blocked by secondary structure or bound proteins like PPR10. This hypothesis obviates the necessity for PPR protein-protein interaction sites, which are vital to the recruitment-based model. In vitro exonuclease assays support this model by demonstrating that recombinant PPR10 can block both 3' → 5' and 5' → 3' exonuclease progression.

Another consequence of PPR10 binding is increased translation of *atpH* RNA. We had previously shown that PPR10 promotes translation of *atpH* mRNAs, although the mechanism was unclear (72). In this study we elucidate this mechanism. We demonstrate

that PPR10 promotes the formation of single-stranded RNA encompassing the ribosome-binding region of the *atpH* transcript. In the absence of PPR10, the Shine-Dalgarno sequence, which is important for ribosomal recruitment, is base paired with part of the PPR10 binding site. We propose that this structure limits ribosomal access to the 5'UTR thereby inhibiting translation. The PPR10/RNA binding site includes the complement sequence to the Shine-Dalgarno site, therefore PPR10 binding prevents formation of this inhibitory structure. This exposes the Shine-Dalgarno and promotes ribosomal recruitment. This model for promoting translation does not rely on PPR protein-protein interactions, but instead rests solely on PPR10s ability to bind a long RNA tract and thereby prevent its base pairing with adjacent sequences.

We believe the mechanism by which PPR5 promotes *trnG* intron splicing mirrors the mechanism by which PPR10 promotes translation. Similar to the case for PPR10, PPR5 binds a stretch of RNA that is adjacent to sequence elements required for splicing: EBS1,  $\delta$ , and  $\alpha'$  (71). All three need to base pair with their complementary sequences for splicing to occur. Without PPR5 present, EBS1, and  $\delta$  are sequestered in a stem loop structure. We propose that PPR5 captures the stem loop sequence in a single-stranded conformation thereby enabling the EBS1, and  $\delta$  elements to base pair with their complementary sequences. In addition, PPR5 induces nuclease hypersensitivity in the  $\alpha'$  sequence suggesting that it augments the structure of this RNA in some way. The significance of this is not clear but it could reflect a conformational change that makes  $\alpha'$  more accessible for base pairing.

We show here that both PPR10 and PPR5 can prevent secondary structure formation by binding single-stranded RNA. We demonstrate how this property results in two disparate PPR functions, ribosomal recruitment in the case of PPR10, and splicing in the case of PPR5. In both cases, the downstream effect can be explained by PPR proteins binding to an extended tract of single-stranded RNA, and thereby influencing the structure of adjacent RNA, instead of directly recruiting effector proteins. We speculate that this mechanism of action may be common among many PPR proteins.

## Future Directions

### Future directions related to WHY1

Despite the many functions that have been attributed to WHY1 and the other whirly family members, there is no clear indication how this family of proteins mediates their downstream effects. The WHY1 phenotype in maize presents as a seedling with virtually no chlorophyll, suggesting chloroplast biogenesis is severely affected. This phenotype can be explained by the near complete loss of ribosomal RNA from the chloroplast. However, how the *why1* mutation leads to rRNA loss is still a mystery that needs to be resolved. We did not find evidence for a direct interaction between WHY1 and ribosomal RNA or ribosomal subunits, but we cannot exclude that WHY1 may influence ribosomal assembly factors. (not really- just brings down the whole nucleoid...)

It is still unresolved whether WHY1 binds to DNA without specificity, or whether it binds a commonly represented sequence. Co-crystallizing WHY1 with DNA would address this issue. The significance of WHY1's DNA binding properties in the chloroplast has yet to be revealed. Though we found no gross defects in DNA amount or nucleoid appearance in *why1* mutant plants, a closer analysis of nucleoid composition and structure could reveal defects suggestive of WHY1 function.

Finally, how WHY1 contributes to *atpF* intron splicing warrants further investigation. Refinement of the WHY1 binding site within the *atpF* intron could give insight into this question.

### Immediate directions related to PPR proteins

Several experiments need to be done to clarify aspects of the work presented in this dissertation. The PPR10 minimal binding site, as determined by alkali hydrolysis binding assays, needs to be validated by GMS assays. We plan to determine whether PPR10 will bind a synthetic RNA oligonucleotide that recapitulates the minimal 3' and 5' ends. The ~15 nt PPR10 binding site is of a manageable size for in depth mutagenesis studies. We will make mutations in this sequence and assay for PPR10 binding ability.

This can give us insight into which, and how many, bases are important for PPR10 interaction. In addition, deleting internal nucleotides can indicate whether PPR10 binds as a ridged structure or whether it has a more flexible binding capacity.

We plan to perform additional exonuclease assays to refine our current findings. We will repeat the PNPase assay under conditions that are more favorable for PNPase activity. This may lead to a stronger correlation between the *in vivo* atpI 3' end, and the 3' end resulting from PPR10 protection against PNPase cleavage. Our results of the 5'→3' terminator nuclease assays could have two explanations: PPR10 blocks terminator nuclease progression, or PPR10 prevents terminator nuclease from binding the RNA. To distinguish between these two possibilities we need to repeat this assay with RNA that contains additional 5' end sequence. This way we can be confident that terminator nuclease loading is unhindered by bound PPR10.

### **Long-term directions related to PPR proteins**

Despite the prevalence and importance for pentatricopeptide repeat proteins in plant organelles, there is relatively little known about the biochemical mechanisms through which they exert downstream effects. This work explores how the putative long RNA interaction surface can mediate several of the functions attributed to PPR proteins. This investigation is one of the few biochemical analyses of this extensive family of proteins. To better understand how PPR proteins mediate downstream effects it will be important to carry out similar analyses with additional members of this protein family. In this way we can determine whether the results presented here are exceptions to the rule, or widely relevant to PPR family members.

PPR5 and PPR10 are the only proteins in this family for which there has been an extensive analysis of the RNA ligand. Identifying and refining additional PPR ligands will allow us to validate several of the predictions made by this study. Our research predicts that many PPR proteins will have similarly long RNA binding sites. It will be interesting to see whether the length of the RNA ligand typically correlates with the number of PPR motifs contained within the protein. Narrowing down the exact binding

sites, and modeling the RNA structures around these sites will provide insight into how specific PPR proteins may influence gene expression. The PPR proteins we have studied mediate downstream effects by promoting the formation of single-stranded RNA. It will be interesting to see whether PPR proteins generally prefer single-stranded RNA as a binding substrate, and if so, whether they tend to bind to sites where secondary structure sequesters sequences important for RNA processing or translation.

Bioinformatic approaches can be used to identify potential PPR binding sites. Many PPR proteins are predicted to bind to intergenic regions of the chloroplast genome. Intergenic regions, in general, have low sequence conservation between plant species. Because PPR proteins bind long RNA tracts with sequence-specificity we predict that PPR binding sites in the intergenic regions will have a higher level of conservation than adjacent sequences. We are currently working with a collaborator, Rodger Voelker, to identify highly conserved sequences within intergenic regions of the chloroplast genome. We believe some of these could correspond to PPR binding sites.

It is still unclear how PPR proteins recognize their RNA ligands. Because PPR motifs are similar to TPR motifs, PPR sequences have been modeled by threading onto known TPR structures. Such models allow us to speculate on what types of interactions are possible between the PPR surface and nucleic acids. These models suggest that PPR proteins may contain asparagine ladders similar to those in ARM repeat proteins. Whereas in ARM repeat proteins these are thought to mediate nonspecific interactions with other proteins, perhaps in PPR proteins they mediate interactions with RNA. Actual crystallographic structures would greatly enhance our understanding of PPR/RNA interactions. Attempts at crystallizing PPR10 with and without its RNA ligand are ongoing with our collaborators, Ian Small and Charles Bond.

The long interaction surface created by the consecutive arrangement of PPR repeats suggests a modular nucleotide interaction motif that can be rearranged to create new binding surfaces specific to new RNA sequences. Perhaps the expanded use of PPR protein in plants arose through gene duplication followed by simple rearrangements or mutations of these repeats. The rearrangements may have resulted in specificity for new

RNA sequences, which were subsequently selected for based on utility. Further understanding of PPR/RNA interactions could enable us to exploit these modules to create *de novo* site-specific RNA interacting proteins.

## REFERENCES

1. Timmis JN, Ayliffe MA, Huang CY, & Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* 5(2):123-135.
2. Kleine T, Maier UG, & Leister D (2008) DNA Transfer From Organelles to the Nucleus: The Idiosyncratic Genetics of Endosymbiosis. *Annu Rev Plant Biol*.
3. Embley TM & Martin W (2006) Eukaryotic evolution, changes and challenges. *Nature* 440(7084):623-630.
4. Allen JF (2003) The function of genomes in bioenergetic organelles. *Philos Trans R Soc Lond B Biol Sci* 358(1429):19-37; discussion 37-18.
5. Maier UG, *et al.* (2008) Complex chloroplast RNA metabolism: just debugging the genetic programme? *BMC Biol* 6:36.
6. Schmitz-Linneweber C & Barkan A (2007) RNA splicing and RNA editing in chloroplasts. *Cell and Molecular Biology of Plastids*, Topics in Current Genetics, ed R B (Springer, Berlin / Heidelberg), Vol 19, pp 213-248.
7. Bollenbach T, Schuster G, Portnoy V, & Stern D (2007) Processing, degradation, and polyadenylation of chloroplast transcripts. *Cell and Molecular Biology of Plastids*, Topics in Current Genetics, ed Bock R (Springer-Verlag, Berlin), pp 175-211.
8. Kroeger T, Watkins K, Friso G, Wijk Kv, & Barkan A (2009) A plant-specific RNA binding domain revealed through analysis of chloroplast group II intron splicing. *Proc Natl Acad Sci U S A* 106:4537-4542.
9. Barkan A, *et al.* (2007) The CRM domain: an RNA binding module derived from an ancient ribosome-associated protein. *RNA* 13:55-64.
10. Small I & Peeters N (2000) The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* 25:46-47.
11. Lurin C, *et al.* (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16(8):2089-2103.



12. Marchfelder A & Binder S (2004) Plastid and plant mitochondrial RNA processing and RNA stability. *Molecular Biology and Biotechnology of Plant Organelles*, eds Daniell H & Chase C (Kluwer Academic Publishers, Dordrecht, The Netherlands), pp 261-294.
13. Nickelsen J (2003) Chloroplast RNA-binding proteins. *Curr Genet* 43(6):392-399.
14. Schwacke R, Fischer K, Ketelsen B, Krupinska K, & Krause K (2007) Comparative survey of plastid and mitochondrial targeting properties of transcription factors in Arabidopsis and rice. *Mol Genet Genomics* 277(6):631-646.
15. Jenkins B, Kulhanek D, & Barkan A (1997) Nuclear mutations that block group II RNA splicing in maize chloroplasts reveal several intron classes with distinct requirements for splicing factors. *Plant Cell* 9:283-296.
16. Jenkins B & Barkan A (2001) Recruitment of a peptidyl-tRNA hydrolase as a facilitator of group II intron splicing in chloroplasts. *EMBO J* 20:872-879.
17. Asakura Y & Barkan A (2007) A CRM domain protein functions dually in group I and group II intron splicing in land plant chloroplasts. *Plant Cell* 19:3864-3875.
18. Asakura Y & Barkan A (2006) Arabidopsis Orthologs of Maize Chloroplast Splicing Factors Promote Splicing of Orthologous and Species-specific Group II Introns. *Plant Physiol* 142:1656-1663.
19. Ostheimer G, *et al.* (2003) Group II intron splicing factors derived by diversification of an ancient RNA binding module. *EMBO J* 22:3919-3929.
20. Till B, Schmitz-Linneweber C, Williams-Carrier R, & Barkan A (2001) CRS1 is a novel group II intron splicing factor that was derived from a domain of ancient origin. *RNA* 7:1227-1238.
21. Watkins K, *et al.* (2007) A ribonuclease III domain protein functions in group II intron splicing in maize chloroplasts. *Plant Cell* 19:2606-2623.
22. Asakura Y, Bayraktar O, & Barkan A (2008) Two CRM protein subfamilies cooperate in the splicing of group IIB introns in chloroplasts. *RNA* 14:2319-2332.
23. Falcon de Longevialle A, *et al.* (2008) The pentatricopeptide repeat gene OTP51 with two LAGLIDADG motifs is required for the cis-splicing of plastid ycf3 intron 2 in Arabidopsis thaliana. *Plant J.* 56:157-168.

24. Schmitz-Linneweber C, *et al.* (2006) A Pentatricopeptide Repeat Protein Facilitates the trans-Splicing of the Maize Chloroplast rps12 Pre-mRNA. *Plant Cell* 18(10):2650-2663.
25. Ostersetzer O, Watkins K, Cooke A, & Barkan A (2005) CRS1, a chloroplast group II intron splicing factor, promotes intron folding through specific interactions with two intron domains. *Plant Cell* 17:241-255.
26. Desveaux D, *et al.* (2004) A "Whirly" transcription factor is required for salicylic acid-dependent disease resistance in Arabidopsis. *Dev Cell* 6(2):229-240.
27. Desveaux D, Despres C, Joyeux A, Subramaniam R, & Brisson N (2000) PBF-2 is a novel single-stranded DNA binding factor implicated in PR-10a gene activation in potato. *Plant Cell* 12(8):1477-1489.
28. Desveaux D, Allard J, Brisson N, & Sygusch J (2002) A new family of plant transcription factors displays a novel ssDNA-binding surface. *Nat Struct Biol* 9(7):512-517.
29. Yoo HH, Kwon C, Lee MM, & Chung IK (2007) Single-stranded DNA binding factor AtWHY1 modulates telomere length homeostasis in Arabidopsis. *Plant J* 49(3):442-451.
30. Krause K, *et al.* (2005) DNA-binding proteins of the Whirly family in Arabidopsis thaliana are targeted to the organelles. *FEBS Lett* 579(17):3707-3712.
31. Pfalz J, Liere K, Kandlbinder A, Dietz KJ, & Oelmuller R (2006) pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell* 18(1):176-197.
32. Marechal A, *et al.* (2008) Overexpression of mtDNA-associated AtWhy2 compromises mitochondrial function. *BMC Plant Biol* 8:42.
33. Williams P & Barkan A (2003) A chloroplast-localized PPR protein required for plastid ribosome accumulation. *Plant J.* 36:675-686.
34. Walbot V & Coe EH (1979) Nuclear gene *iojap* conditions a programmed change to ribosome-less plastids in *Zea mays*. *Proc. Natl. Acad. Sci. USA* 76:2760-2764.
35. Barkan A (1993) Nuclear mutants of maize with defects in chloroplast polysome assembly have altered chloroplast RNA metabolism. *Plant Cell* 5:389-402.
36. Barkan A (1998) Approaches to investigating nuclear genes that function in chloroplast biogenesis in land plants. *Methods Enzymol.* 297:38-57.

37. Voelker R & Barkan A (1995) Nuclear genes required for post-translational steps in the biogenesis of the chloroplast cytochrome *b6f* complex. *Molec. Gen. Genet.* 249:507-514.
38. Barkan A (2008) Genome-wide analysis of RNA-protein interactions in plants. *Plant Systems Biology, Methods in Molecular Biology*, ed Belostotsky D (Humana Press).
39. Voelker R, Mendel-Hartvig J, & Barkan A (1997) Transposon-disruption of a maize nuclear gene, *tha1*, encoding a chloroplast SecA homolog: *in vivo* role of cp-SecA in thylakoid protein targeting. *Genetics* 145:467-478.
40. Wong I & Lohman T (1993) A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proc. Natl. Acad. Sci., U.S.A.* 90:5428-5432.
41. Mullet J & Klein R (1987) Transcription and RNA stability are important determinants of higher plant chloroplast RNA levels. *EMBO J* 6:1571-1579.
42. Rapp JC, Baumgartner BJ, & Mullet J (1992) Quantitative analysis of transcription and RNA levels of 15 barley chloroplast genes: transcription rates and mRNA levels vary over 300-fold; predicted mRNA stabilities vary 30-fold. *J. Biol. Chem.* 267:21404-21411.
43. Klein R & Mullet J (1990) Light-induced transcription of chloroplast genes. *J. Biol. Chem.* 265:1895-1902.
44. Emanuelsson O & Heijne Gv (2001) Prediction of organellar targeting signals. *Biochim Biophys Acta* 1541:114-119.
45. Small I, Peeters N, Legeai F, & Lurin C (2004) Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* 4(6):1581-1590.
46. Sato N, Terasawa K, Miyajima K, & Kabeya Y (2003) Organization, developmental dynamics, and evolution of plastid nucleoids. *Int Rev Cytol* 232:217-262.
47. Schmitz-Linneweber C, Williams-Carrier R, & Barkan A (2005) RNA immunoprecipitation and microarray analysis show a chloroplast pentatricopeptide repeat protein to be associated with the 5'-region of mRNAs whose translation it activates. *Plant Cell* 17:2791-2804.

48. Beick S, Schmitz-Linneweber C, Williams-Carrier R, Jensen B, & Barkan A (2008) The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. *Mol. Cell. Biol.* 28:5337-5347.
49. Hess WR, *et al.* (1994) Inefficient *rpl2* splicing in barley mutants with ribosome-deficient plastids. *Plant Cell* 6:1455-1465.
50. Vogel J, Boerner T, & Hess W (1999) Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants. *Nucl. Acids Res.* 27:3866-3874.
51. Barkan A (1989) Tissue-dependent plastid RNA splicing in maize: Transcripts from four plastid genes are predominantly unspliced in leaf meristems and roots. *Plant Cell* 1:437-445.
52. Stahl DJ, Rodermeil SR, Bogorad L, & Subramanian AR (1993) Co-transcription pattern of an introgressed operon in the maize chloroplast genome comprising four ATP synthase subunit genes and the ribosomal *rps2*. *Plant Molec. Biol.* 21:1069-1076.
53. Schumacher MA, Karamooz E, Zikova A, Trantirek L, & Lukes J (2006) Crystal structures of *T. brucei* MRP1/MRP2 guide-RNA binding complex reveal RNA matchmaking mechanism. *Cell* 126(4):701-711.
54. Bollenbach TJ, *et al.* (2005) RNR1, a 3'-5' exoribonuclease belonging to the RNR superfamily, catalyzes 3' maturation of chloroplast ribosomal RNAs in *Arabidopsis thaliana*. *Nucleic Acids Res* 33(8):2751-2763.
55. Bellaoui M, Keddie JS, & Gruissem W (2003) DCL is a plant-specific protein required for plastid ribosomal RNA processing and embryo development. *Plant Mol Biol* 53(4):531-543.
56. Bellaoui M & Gruissem W (2004) Altered expression of the *Arabidopsis* ortholog of DCL affects normal plant development. *Planta* 219(5):819-826.
57. Bisanz C, *et al.* (2003) The *Arabidopsis* nuclear DAL gene encodes a chloroplast protein which is required for the maturation of the plastid ribosomal RNAs and is essential for chloroplast differentiation. *Plant Mol Biol* 51(5):651-663.
58. Zaegel V, *et al.* (2006) The plant-specific ssDNA binding protein OSB1 is involved in the stoichiometric transmission of mitochondrial DNA in *Arabidopsis*. *Plant Cell* 18(12):3548-3563.
59. Dorman CJ & Deighan P (2003) Regulation of gene expression by histone-like proteins in bacteria. *Curr Opin Genet Dev* 13(2):179-184.

60. Kamashev D, Balandina A, Mazur AK, Arimondo PB, & Rouviere-Yaniv J (2008) HU binds and folds single-stranded DNA. *Nucleic Acids Res* 36(3):1026-1036.
61. Kobayashi T, *et al.* (2002) Detection and localization of a chloroplast-encoded HU-like protein that organizes chloroplast nucleoids. *Plant Cell* 14(7):1579-1589.
62. Sato N (2001) Was the evolution of plastid genetic machinery discontinuous? *Trends Plant Sci* 6(4):151-155.
63. Sato N, Nakayama M, & Hase T (2001) The 70-kDa major DNA-compacting protein of the chloroplast nucleoid is sulfite reductase. *FEBS Lett* 487(3):347-350.
64. Sekine K, *et al.* (2007) DNA binding and partial nucleoid localization of the chloroplast stromal enzyme ferredoxin:sulfite reductase. *Febs J* 274(8):2054-2069.
65. Lia G, *et al.* (2003) Supercoiling and denaturation in Gal repressor/heat unstable nucleoid protein (HU)-mediated DNA looping. *Proc Natl Acad Sci U S A* 100(20):11373-11377.
66. Kar S, Edgar R, & Adhya S (2005) Nucleoid remodeling by an altered HU protein: reorganization of the transcription program. *Proc Natl Acad Sci U S A* 102(45):16397-16402.
67. Lewis DE, Geanacopoulos M, & Adhya S (1999) Role of HU and DNA supercoiling in transcription repression: specialized nucleoprotein repression complex at gal promoters in *Escherichia coli*. *Mol Microbiol* 31(2):451-461.
68. Balandina A, Kamashev D, & Rouviere-Yaniv J (2002) The bacterial histone-like protein HU specifically recognizes similar structures in all nucleic acids. DNA, RNA, and their hybrids. *J Biol Chem* 277(31):27622-27628.
69. Balandina A, Claret L, Hengge-Aronis R, & Rouviere-Yaniv J (2001) The *Escherichia coli* histone-like protein HU regulates rpoS translation. *Mol Microbiol* 39(4):1069-1079.
70. Schmitz-Linneweber C & Small I (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci* 13(12):663-670.
71. Williams-Carrier R, Kroeger T, & Barkan A (2008) Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. *RNA* 14:1930-1941.

72. Pfalz J, Bayraktar O, Prikryl J, & Barkan A (2009) Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO J* 28:2042-2052.
73. Okuda K, Nakamura T, Sugita M, Shimizu T, & Shikanai T (2006) A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J Biol Chem* 281(49):37661-37667.
74. Hayes R, *et al.* (1996) Chloroplast mRNA 3'-end processing by a high molecular weight protein complex is regulated by nuclear encoded RNA binding proteins. *EMBO J* 15:1132-1141.
75. Tsuchiya N, Fukuda H, Sugimura T, Nagao M, & Nakagama H (2002) LRP130, a protein containing nine pentatricopeptide repeat motifs, interacts with a single-stranded cytosine-rich sequence of mouse hypervariable minisatellite Pc-1. *Eur. J. Biochem.* 269:2927-2933.
76. Nakamura T, Meierhoff K, Westhoff P, & Schuster G (2003) RNA-binding properties of HCF152, an Arabidopsis PPR protein involved in the processing of chloroplast RNA. *Eur J Biochem* 270(20):4070-4081.
77. Steitz JA & Jakes K (1975) How ribosomes select initiator regions in mRNA: base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in Escherichia coli. *Proc Natl Acad Sci U S A* 72(12):4734-4738.
78. Pyle A & Lambowitz A (2006) Group II introns: ribozymes that splice RNA and invade DNA. *The RNA World*, eds Gesteland R, Cech T, & Atkins J (Cold Spring Harbor Press), 3rd Ed, pp 469-506.
79. Menger M, Tuschl T, Eckstein F, & Porschke D (1996) Mg(2+)-dependent conformational changes in the hammerhead ribozyme. (Translated from eng) *Biochemistry* 35(47):14710-14716 (in eng).
80. Glisovic T, Bachorik JL, Yong J, & Dreyfuss G (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett* 582(14):1977-1986.
81. Lunde BM, Moore C, & Varani G (2007) RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* 8(6):479-490.
82. Auweter SD, Oberstrass FC, & Allain FH (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res* 34(17):4943-4959.

83. Kajava AV (2001) Review: proteins with repeated sequence--structural prediction and modeling. *J Struct Biol* 134(2-3):132-144.
84. D'Andrea LD & Regan L (2003) TPR proteins: the versatile helix. *Trends Biochem Sci* 28(12):655-662.
85. Jinek M, *et al.* (2004) The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. *Nat Struct Mol Biol* 11(10):1001-1007.
86. Wharton RP & Aggarwal AK (2006) mRNA regulation by Puf domain proteins. *Sci STKE* 2006(354):pe37.
87. Wang X, McLachlan J, Zamore PD, & Hall TM (2002) Modular recognition of RNA by a human pumilio-homology domain. *Cell* 110(4):501-512.
88. Loiselay C, *et al.* (2008) Molecular identification and function of cis- and trans-acting determinants for petA transcript stability in *Chlamydomonas reinhardtii* chloroplasts. *Mol Cell Biol* 28(17):5529-5542.
89. Hattori M & Sugita M (2009) A moss pentatricopeptide repeat protein binds to the 3' end of plastid clpP pre-mRNA and assists with mRNA maturation. *FEBS J* 276(20):5860-5869.